

JÉSSICA COPETTI

**AS MÚLTIPLAS PERDAS DA CIS-ACONITATO
DESCARBOXILASE AO LONGO DA EVOLUÇÃO DOS
METAZOÁRIOS ESTÃO RELACIONADAS AO EFEITO
TÓXICO DE SEU PRODUTO ENZIMÁTICO**

Trabalho de Conclusão de Curso
submetido à Universidade Federal de
Santa Catarina para a obtenção do grau de
bacharel em Ciências Biológicas.
Orientador: Prof. Dr. Daniel Santos
Mansur.

Florianópolis
2018

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Copetti, Jéssica

As múltiplas perdas da cis-aconitato
descarboxilase ao longo da evolução dos metazoários
estão relacionadas ao efeito tóxico de seu produto
enzimático / Jéssica Copetti ; orientador, Daniel
Santos Mansur, 2018.

93 p.

Trabalho de Conclusão de Curso (graduação) -
Universidade Federal de Santa Catarina, Centro de
Ciências Biológicas, Graduação em Ciências Biológicas,
Florianópolis, 2018.

Inclui referências.

1. Ciências Biológicas. 2. evolução molecular,
imunometabolismo, bioinformática. I. Santos Mansur,
Daniel . II. Universidade Federal de Santa
Catarina. Graduação em Ciências Biológicas. III.
Título.

AGRADECIMENTOS

Inicialmente agradeço às agências de fomento à pesquisa, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), que disponibilizaram recursos para a execução deste trabalho e de parte de minha experiência enquanto estudante em iniciação científica.

Agradeço igualmente a oportunidade de poder completar meus estudos em ciências biológicas na Universidade Federal de Santa Catarina com total isenção de custos. Compadeço-me de todas/os as/os demais brasileiras/os que, diferente de mim, não podem sonhar este sonho.

A partir daqui demonstrarei meu agradecimento às pessoas que tiveram papéis especiais durante meu percurso acadêmico e que ao cederem um tanto de si, transformaram minha mente e fizeram de mim uma aprendiz à cientista e uma criatura feliz.

Agradeço imensamente ao Daniel Mansur, meu orientador e amigo por quatro anos e meu principal pai científico. Sem suas ideias este trabalho não existiria e sem seu apoio incondicional eu provavelmente estaria vendendo telas.

Agradeço à altura ao André Báfica, padrinho científico, o não-biólogo mais biólogo que conheço! Sua audácia argumentativa é infecciosa. Gratidão por todas as minhocas em minha cabeça.

Agradeço às/aos colegas do Laboratório de Imunobiologia pela transmissão de experiência e suporte de todos os dias e, claro, pelos cafés. Agradeço em especial ao Edgar, que me ajudou construir a base de meu conhecimento de programação e que me socorreu várias vezes no terminal. Sem você este trabalho não teria figuras! *Muchas gracias!*

Agradeço ao José (Zé) Henrique de Oliveira pela ajuda com os experimentos preliminares de toxicidade do ácido itacônico, assim como por sua presença radiante. Agradeço igualmente seu aceite para participar da banca examinadora deste trabalho.

Agradeço de antemão as demais integrantes da banca, Luisa Rona e Gabriela Luiz, por terem aceite de prontidão o meu convite de participar desse marco em minha trajetória acadêmica. Um viva às mulheres cientistas!

Agradeço com um abraço apertadíssimo às minhas mães científicas Michele, Alice, Adara e Constanza. Salve, rainhas! Acompanharam-me atenciosamente na bancada, me ensinaram as manhas e os segredinhos, me deram as broncas que eu precisava ouvir e acima de tudo, tornaram-se minhas amigas queridas.

Agradeço profundamente ao Kay Saalfeld por me ajudar a desmascarar a biologia evolutiva travestida de religião que os acadêmicos papagaiam. Meu percurso acadêmico foi marcado por três principais *inputs*: a experiência em sala/campo, a iniciação científica nos laboratórios e os encontros do grupo de evolução. Não esqueço, é claro, das/os amigas/os que construíram o grupo coletivamente, em especial André, Ângelo, Beatriz, Bruna, Dani, Hugo, João, Joseane, Kauan, Murilo e Tomás, e mesmo que não mencionados, sintam-se incluídos e agradecidos todos os demais participantes. *With a little help from my friends*, entendi a evolução como contingência histórica e só a partir daí a transmutei em verdadeiro campo investigativo.

Agradeço às/aos professoras/es que tive a oportunidade de conhecer durante a graduação e que compartilharam seus conhecimentos e vidas conosco, as/os estudantes. Agradeço em especial ao Giordano Calloni, meu primeiro orientador de iniciação científica, e à Luciane Perazzolo, que me apresentaram o fantástico mundo intracelular. Desde então meu encantamento pelo microscópico se desenovelou.

Agradeço com um abraço coletivo imenso às amigas e amigos do curso de biologia. Recebi tanto e de tantos! Eu não conseguiria agradecer pelo tudo que se somou a mim durante sete curtíssimos anos ao lado de vocês. Vocês mantêm a colorida utopia viva em mim. Agradeço especialíssimamente às minhas amigas queridas Kelly e Lary, assim como à Manu, Vanessa e Marcelo, pela parceria ao longo dos anos, pelas risadas e comidinhas.

Por fim, e não menos importante, agradeço aos que mantiveram minha sanidade mental quando precisei: à minha mãezinha, poço de bondade, pelo amor sem limites desde o início de minha existência e ao meu paizinho, guia da minha força de vontade, onde quer que esteja. Agradeço igualmente o carinho dos demais integrantes de minha família, especialmente à minha sobrinha fofíssima Laura. Acrescento aqui um pensamento especial direcionado a meu mestre yogi Fernando. Sem sua luz o caminho teria sido mais escuro.

À todas/os, aprendi muito mais que apenas biologia com vocês.
Admiro e amo a todas/os!
Gratidão! Ahó!

RESUMO

Em resposta ao contato com microrganismos ou situações inflamatórias, as células fagocitárias dos metazoários mobilizam centenas de moléculas: citocinas, quimiocinas, moléculas de adesão e muitas outras. Entre elas, o ácido itacônico. Este composto é produto enzimático da cis-aconitato descarboxilase e exerce atividade antimicrobiana sobre bactérias patogênicas intracelulares nos vertebrados. As proteínas homólogas à cis-aconitato descarboxilase estão espalhadas por toda a árvore da vida, entretanto, muitas espécies a perderam. Ao investigarmos as causas das múltiplas deleções, identificamos uma forte correlação entre as perdas gênicas e a ausência de uma via de degradação do ácido itacônico. Sugerimos, portanto, que uma possível causa das perdas gênicas que moldaram a distribuição desse gene nas espécies foi a pressão seletiva negativa sobre os indivíduos sem meios de detoxificação.

Palavras-chave: *Immune responsive gene 1*. Ácido itacônico. Itaconato. Citramalil-coenzima A liase. Seleção natural.

ABSTRACT

Facing microorganisms or inflammatory conditions, metazoan phagocytic cells mobilize hundreds of molecules: cytokines, chemokines, adhesion molecules and many others. Among them is the itaconic acid. This compound is the product of the cis-aconitate decarboxylase enzyme and acts as an antimicrobial metabolite, killing intracellular pathogenic bacteria. Protein homologues of cis-aconitate decarboxylase are widely spread along the tree of life, however, many species have undergone gene losses. We investigated the causes underlying cis-aconitate gene losses and we found out a strong correlation between these events and the lack of an itaconic acid degradation pathway. Thus, we suggest that a plausible cause for the gene losses observed, which shaped cis-aconitate decarboxylase distribution among species, is a negative selective pressure on individuals lacking detoxifying means.

Keywords: *Immune responsive gene 1*. Itaconic acid. Itaconate. Citramalyl-coenzyme A lyase. Natural selection.

SUMÁRIO

1 INTRODUÇÃO.....	10
1.1 Biologia evolutiva.....	10
1.1.1 A origem da vida.....	10
1.1.2 Seleção natural.....	11
1.1.3 Hereditariedade.....	11
1.1.4 Origem da variação.....	12
1.1.5 Genética evolutiva.....	14
1.2 Sistemas imunológicos inatos.....	14
1.2.1 Origem evolutiva dos sistemas imunológicos.....	14
1.2.2 Receptores de reconhecimento de padrão molecular.....	17
1.2.3 Sinalização celular e sistemas efetores imunobiológicos...18	
1.2.4 A enzima cis-aconitato descarboxilase e seu produto, o ácido itacônico.....	21
1.3 Pergunta.....	24
1.4 Objetivos.....	27
1.4.1 Objetivo geral	27
1.4.2 Objetivos específicos.....	27
2 METODOLOGIA.....	28
2.1 Árvores filogenéticas.....	28
2.1.1 Obtenção e preparação das sequências.....	28
2.1.2 Alinhamento e geração de árvores.....	28
2.2 Cladogramas.....	30
2.3 Heatmap.....	30
2.4 Nomenclatura taxonômica.....	30
2.5 Limitações da pesquisa.....	31
3 RESULTADOS E DISCUSSÃO.....	33
3.1 As proteínas da família MmgE/PrpD estão presentes nos três domínios da vida.....	33
3.2 As proteínas da família MmgE/PrpD constituem duas subfamílias distintas.....	38
3.3 Moluscos, cefalocordados e vertebrados são os únicos filos de Metazoa em que consta a presença da cis-aconitato descarboxilase.41	
3.4 As cis-aconitato descarboxilases parálogas encontradas nos metazoários aparentemente mantêm a mesma função.....	43

3.5 O gene codificador da cis-aconitato descarboxilase foi perdido múltiplas vezes ao longo da evolução.....	45
3.6 O ácido itacônico tem efeito tóxico.....	48
3.6.1 A citramalil-coenzima A liase é provavelmente ubíqua no processo de degradação do ácido itacônico.....	51
3.6.2 Todos os metazoários portadores de cis-aconitato descarboxilase possuem também citramalil-coenzima A liases.....	52
3.7 A ausência de um mecanismo de controle negativo da produção de ácido itacônico torna a presença da cis-aconitato descarboxilase incompatível com a vida dos metazoários.....	55
3.8 A evolução da cis-aconitato descarboxilase nos vertebrados é multifatorial.....	57
3.8.1 A ausência do Immune responsive gene 1 durante o processo de implantação embrionário em camundongos implica em taxas abortivas de 50-100 %.....	58
3.8.2 Coevolução de microrganismos intracelulares e hospedeiros pode ter pressionado a manutenção da cis-aconitato descarboxilase em moluscos e vertebrados.....	59
3.9 Conclusões.....	60
REFERÊNCIAS.....	63
APÊNDICES.....	82

1 INTRODUÇÃO

1.1 BIOLOGIA EVOLUTIVA

1.1.1 A origem da vida

A vida na Terra surgiu há cerca de quatro bilhões de anos. Nesse momento, a superfície do planeta estava tomada por um vasto oceano e a atmosfera tropical era aquecida por radiação intensa (ARNDT; NISBET, 2012).

Os sistemas precursores dos seres vivos foram redes metabólicas baseadas em enxofre ou ácidos ribonucleicos (RNAs, *ribonucleic acids*) autorreplicantes que provavelmente emergiram em fontes hidrotermais, escapes vulcânicos da crosta oceânica (ARNDT; NISBET, 2012; ORGEL, 2004; RUSSEL; DANIEL; HALL, 1993). Fontes como essas continuam abundantes no leito oceânico e a partir delas, vários compostos orgânicos simples são transferidos à hidrosfera (MARTIN et al., 2008).

Em uma era pré-membranar, as protobiomoléculas formadoras dos seres vivos se difundiriam livremente pelo oceano e não coexistiriam em concentrações suficientes para que interagissem se não houvesse um mecanismo de contenção (RUSSEL; HALL, 1997). Acredita-se que os poros inorgânicos das fontes hidrotermais submarinas, com dimensões celulares, tenham sido os primeiros compartimentos a circundar as redes moleculares nascentes (BAASKE et al., 2007; RUSSEL et al., 1994; RUSSEL; HALL, 1997). As cooperativas autorreplicantes que comporiam as primeiras células se perpetuaram nesses poros, os quais mantinham a condição redutora e a ciclagem de temperaturas favoráveis às interações moleculares (BROWN; GODDARD, 2003).

Os eventos independentes de substituição dos envoltórios minerais por membranas lipídicas nos ancestrais bacterianos e arqueanos, assim como a consolidação de interações bioquímicas mais sofisticadas, concluiu a libertação e surgimento das protocélulas (LOMBARD; LÓPEZ-GARCÍA; MOREIRA, 2012).

1.1.2 Seleção natural

A estabilidade e a replicabilidade, equivalentes à sobrevivência e reprodução nos seres vivos, foram os primeiros fenótipos das estruturas pré-bióticas a passarem por um filtro seletivo (DE DUVE, 2005). O processo adaptivo de entidades em transformação, tal como a sopa de moléculas pré-bióticas, ocorre por meio de eventos de seleção sobre a variação no fenótipo dos indivíduos em interação com o ambiente. A construção fenotípica inicia com a história de vida da matriz, tudo o que se desenrola durante o desenvolvimento contribui para a variação fenotípica manifestada e durante todo o tempo de existência o indivíduo está suscetível à mudanças e às pressões seletivas sobre as variações produzidas. As moléculas pré-bióticas que sob as condições ambientais da Terra hadeana permaneceram relativamente estáveis enquanto entidades sobreviveram.

As populações biológicas passam por um processo relativamente contínuo de transformação, em que o surgimento de variações (próximo tópico, 1.1.4), se não-deletérias, é seguido por eventuais afunilamentos populacionais, promovidos por catástrofes naturais, fluxos migratórios, epidemias e outros fenômenos perturbadores da arquitetura populacional, que culminam na limitação da diversidade fenotípica (DARWIN, 1869; LANDE, 1988; ZUCKERKANDL; PAULING, 1965). O efeito fundador que se segue, ao longo dos anos seguintes ao gargalo populacional, se caracteriza pela hibridização entre os sobreviventes, os quais compreendem diversidade reduzida em relação à população da geração anterior, e assim concluem a disseminação das características fenotípicas dos indivíduos sobreviventes de forma equivalente na população (LANDE, 1988; VORZIMMER, 1963).

1.1.3 Hereditariedade

Sem nenhuma exceção conhecida, todas as células que conhecemos atualmente armazenam informações desenvolvimentais em forma de moléculas de ácido desóxirribonucleico (DNA, *deoxiribonucleic acid*), fitas bioquímicas entrelaçadas em hélice, não ramificadas e em geral compostas por quatro monômeros: adenina, citosina, timina e guanina, ACTG (ALBERTS et al., 2010, p.2). Além

do código genético, os sistemas de transcrição e tradução, a quiralidade dos açúcares e aminoácidos¹ e as membranas lipídicas (apesar de suas naturezas moleculares distintas) são características compartilhadas por todos os seres vivos, o que nos motiva a afirmar que os indivíduos que compunham a população ancestral da vida na Terra usufruíam dessas mesmas características (KOONIN; MARTIN, 2005; LOMBARD; LÓPEZ-GARCÍA; MOREIRA, 2012).

O DNA, acima dos demais componentes do fenótipo, recebe interesse especial devido à sua propriedade intrínseca de transmissão de informação ao longo de gerações de reprodução. O DNA porta diretrizes sobre a arquitetura proteica da célula e sua regulação, a partir das quais, somadas às condições pré-estabelecidas do microambiente de nascimento e aos estímulos percebidos durante o desenvolvimento, o fenótipo da espécie é reconstruído em cada indivíduo (DISCHER; MOONEY, ZANDSTRA, 2009; KUMAR; PLACONE; ENGLER, 2017).

O genótipo da espécie é a consequência secundária do sucesso reprodutivo do conjunto de fenótipos produzidos nos diferentes estágios de vida dos organismos e em diferentes condições ambientais e, devido à sua propriedade hereditária, confere memória de longo prazo à experiência evolutiva (WEST-EBERHARD, 2003, p. 16). Por meio da análise de sua mudança ao longo das gerações, os cientistas aprendem sobre a resposta das espécies à dinâmica espaço-temporal a que foram submetidas no passado.

1.1.4 Origem da variação

A construção do fenótipo não pode ser atribuída individualmente ao DNA, afinal ele age de forma intrincada junto aos ambientes intra- e extracelular (FELIX, 2016). A comunicação entre o arcabouço gênico e o ambiente durante o desenvolvimento é a força-motriz que produz variações entre os indivíduos.

Há três principais fontes de variações que podem ser identificadas nesse processo. A primeira é a clássica variação genética decorrente de

¹ Utilização exclusiva de apenas um dos estereoisômeros de açúcares e aminoácidos para funções bioquímicas fundamentais: L-aminoácidos e D-açúcares (HARRIS et al., 2003; KOONIN, 2000).

mutações: deleções, substituições, inserções, recombinações e outros processos moleculares. Apesar de (geralmente e relativamente) randômicas, mutações não produzem variações fenotípicas igualmente randômicas (SEGOVIA; TAM; STIRLING, 2015). Alterações específicas ocorrem nas propriedades físicas dos indivíduos conforme o direcionamento da nova instrução gênica. Essa anisotropia é consequência direta do envolvimento do processo desenvolvimental na construção do fenótipo, referido algumas vezes como “viés desenvolvimental” (FELIX, 2016; LALAND et al., 2015).

A segunda principal fonte de variação é decorrente de variações microambientais estocásticas. “Ruído” é intrínseco aos processos moleculares e em alguns casos determinante do fenótipo, no caso de trajetórias desenvolvimentais sensíveis. Por exemplo, em *Caenorhabditis elegans*, o destino da diferenciação de células gonadais é definido por ondas de *feedback* positivo sobre a transcrição gênica, que amplificam as condições estocásticas iniciais (WILKINSON; FITZGERALD; GREENWALD, 1994).

A terceira e última principal fonte de variação nos organismos é fruto de variações parentais que se manifestam na geração seguinte. Por exemplo, na planta *Arabidopsis thaliana*, a temperatura a que a planta mãe é submetida durante o período reprodutivo influencia diretamente a condição de dormência das plantas filhas devido à deposição diferencial de tecido silicoso e tanino sobre as sementes (CHEN et al., 2014). Fenômenos biológicos desse tipo também são referenciados como “plasticidade tardia”, por afetarem apenas a geração seguinte.

Todos os mecanismos causais atuantes sobre a reconstrução do fenótipo na prole devem ser considerados sob a perspectiva evolutiva e os mencionados não são os únicos. Existem outras formas de herança “não-clássicas” que podem ser consideradas aqui, como a transmissão de cultura entre indivíduos, que certamente contribui para a bagagem hereditária da espécie ou grupo e promove a disseminação de inovações (LALAND et al., 2015). Falta-nos entender a fisiologia da hereditariedade de outros processos já bem estabelecidos na ciência, como a epigenética e a plasticidade fenotípica, que, mesmo que obscuras, conduzem a geração de fenótipos diferenciados (FELIX, 2016).

1.1.5 Genética evolutiva

A partir da união das noções de hereditariedade e evolução das espécies, nasce o campo da genética evolutiva. Juntas, as redes de interação molecular e as variações ambientais constroem os fenótipos dos indivíduos por intermédio do desenvolvimento (FELIX, 2016). Dessa forma, há, portanto, vias moleculares, nucleicas e proteicas, correspondentes à expressão e/ou regulação das funções ou módulos fenotípicos manifestados nas espécies.

A detecção de similaridade entre sequências de diferentes espécies permite a inferência de relacionamentos evolutivos, assim como a extrapolação de funções fenotípicas (FANG et al., 2010; KOONIN, 2015). O relacionamento entre as sequências geralmente é examinado ao ordená-las por métodos comparativos de distância, os quais resultam em um diagrama em forma de árvore (YANG; RANNALA, 2012 para uma revisão sobre o tema). Árvores filogenéticas providenciam um método comparativo poderoso para identificar em princípio qualquer processo evolutivo, sendo os mais recorrentes especiação, duplicação, fissão/fusão, deleção e transferência horizontal. A partir da inferência de uma árvore, as causas e consequências dos eventos evolutivos detectados podem ser abordadas (FANG et al., 2010).

A arquitetura gênica e a ecologia das populações se complementam ao interferir na sobrevivência das espécies, logo, as conclusões tiradas a partir de análises genômicas são em seguida sobrepostas aos dados referentes à história de vida dos respectivos organismos, o que permite, finalmente, a construção de um panorama evolutivo do ácido nucleico/proteína, espécie ou fenótipo alvo (KELLER; WALLER, 2002; SACCHERI; HANSKI, 2006).

1.2 SISTEMAS IMUNOLÓGICOS INATOS

1.2.1 Origem evolutiva dos sistemas imunológicos

Os seres vivos pioneiros eram procariotos unicelulares quimio-ou fotolitotróficos que viviam como células livres e cujas atividades se restringiam à absorção e conversão de formas variadas de carbono a

elementos fisiológicos, estruturais e energéticos e, possivelmente, alguma competição por recursos locais (BENGTSON, 2002). A emergência de infecções por entidades virais, independentemente de quando tenham surgido – simultânea ou tardiamente às células –, forçou essas populações ao desenvolvimento de estratégias adaptativas de contenção (HOLMES, 2011).

Estratégias imunológicas têm como um de seus resultados a destruição dos agentes infectantes. O desenvolvimento de mecanismos de destruição nas espécies exigiu que previamente houvessem formas de discriminação entre os componentes próprios e os não-próprios (ambiente, alimento, predadores e invasores), afinal, a eliminação de elementos próprios tende a ser letal ou debilitante para o indivíduo (COHN, 2010). A delimitação da consciência estrutural dos indivíduos unicelulares constituiu a base não apenas para o desenvolvimento de estratégias imunológicas e de nutrição, mas igualmente previne o canibalismo intraespecífico, assim como permite o reconhecimento de compatibilidade entre seres sexuados, que se desenvolveriam no futuro (DZIK, 2010; WADDELL; DUFFY, 1986). Sendo assim, a distinção entre elementos próprios e não-próprios é determinante para eventos de especiação.

De forma geral, a resposta a agentes externos depende de sensores químicos e da transferência da informação obtida através deles a efetores de atuação altamente específica, em conformidade com a natureza do estímulo inicial (DZIK, 2010). O agente infectante interage com receptores, solúveis ou ligados às membranas lipídicas e paredes celulares, os quais são capazes de reconhecer padrões moleculares relativamente invariáveis (JANEWAY, 1989; MEDZHITOV; JANEWAY, 2000). Essa interação resulta na mobilização de mecanismos imunobiológicos efetores que, eventualmente eliminam os agentes patogênicos do hospedeiro.

Nas bactérias e arqueas, após o estabelecimento de um mecanismo de reconhecimento pessoal, enzimas de restrição e outros mecanismos de controle de infecções virais foram desenvolvidos, por exemplo, o sistema CRISPR (abreviação para *Clustered Regularly Interspaced Short Palindromic Repeats*), cujo locus está presente em 45 % das bactérias e 84 % das arqueas (BARRANGOU et al., 2007; GRISSA; VERGNAUD; POURCEL, 2007; JANSEN et al., 2002). Este

mecanismo se baseia na leitura e clivagem de repetições palindrômicas dos genomas dos vírus infectantes e resulta na inativação viral, seguida de estabelecimento de uma memória imunológica gênica, o que põe em cheque a ideia de imunidade adaptativa exclusiva de vertebrados (BARRANGOU et al., 2007; JANSEN et al., 2002).

Da mesma maneira, as primeiras células eucarióticas e seus ancestrais desenvolveram uma série de estratégias de manutenção de sua integridade celular e sobrevivência. Uma delas, determinante para o meio de vida da grande maioria dos seres eucariotos derivados é a fagocitose (METCHNIKOFF, 1905).

De forma simplificada, a fagocitose é a habilidade de capturar elementos do meio via receptores celulares. Protistas ameboides fagocitam como forma de alimentação e de defesa por meio de um mecanismo semelhante ao dos bem estudados macrófagos de mamíferos: reconhecem agentes externos via receptores de superfície celular e os desmontam com radicais livres de oxigênio (ALLEN; DAWIDOWICZ, 1990; DAVIES; CHATTINGS; EDWARDS, 1991). Coanoflagelados, por sua vez, que são protistas unicelulares coloniais que constituem o grupo-irmão dos metazoários, possuem uma série de proteínas envolvidas na adesão e sinalização celular típicas do processo fagocítico de metazoários, como lectinas, caderinas e tirosina quinases o que é um indício de que as bases de vias imunobiológicas fundamentais estavam fundadas antes da consolidação do grupo Metazoa (KING; HITTINGER; CARROLL, 2003).

Fica evidente aqui a artificialidade das divisões conceituais biológicas. O “sistema imunológico” não têm fronteiras bem delimitadas no indivíduo. Seus mecanismos surgiram como redes de interação que desempenham funções tão diversas quanto reconhecimento pessoal, nutrição, locomoção e proliferação. Os conceitos surgem apenas após o desenvolvimento da linguagem humana, ineficiente ao descrever fielmente processos biológicos complexos.

Todos os animais possuem células livres com características fagocíticas (HARTENSTEIN, 2006). Em animais acelomados e que, portanto, não possuem um sistema suporte de transporte fluidoico, como os integrantes dos filos Cycliophora, Entoprocta, Gastrotricha, Gnathostomulida e Platyhelminthes, essas células recebem as denominações “células intersticiais” ou “amebócitos”, enquanto em

animais celomados recebem geralmente o nome de “hemócitos”, embora sejam evolutivamente equivalentes (HARTENSTEIN, 2006; BRUSCA; BRUSCA, 2003). Nos invertebrados, de forma geral, as células fagocíticas participam de processos desenvolvimentais e metabólicos, além de executarem as funções básicas de nutrição e eliminação de agentes infecciosos (DZIK, 2010; METCHNIKOFF, 1905). Nos vertebrados, as células fagocitárias incorporam recursos nutritivos, reciclam moléculas durante a embriogênese, mantêm a homeostase dos tecidos no adulto, eliminam agentes infecciosos e, após surgimento dos complexos principais de histocompatibilidade (MHCs, *major histocompatibility complex*), a habilidade fagocítica foi integrada à imunidade adaptativa dos vertebrados (DESJARDINS; HOUDE; GAGNON, 2005).

1.2.2 Receptores de reconhecimento de padrão molecular

As células fagocitárias são células mediadoras do sistema imunológico inato que reconhecem moléculas do hospedeiro e moléculas exógenas via receptores de reconhecimento de padrão (PRRs, *pattern-recognition receptors*) e que, a partir dessa interação, induzem a expressão de citocinas e outras moléculas que ativam mecanismos de defesa antimicrobiana (JANEWAY, 1989).

Os PRRs são estruturas especializadas que reconhecem elementos ausentes no hospedeiro e amplamente distribuídas nos microrganismos, as quais são supostamente essenciais para a sobrevivência microbiana, o que faz dos PRRs uma estratégia adaptativa de alta eficiência, pois elementos biológicos essenciais têm baixa tolerância à alteração estrutural (MEDZHITOV, JANEWAY, 2000). Os PRRs são proteínas heterogêneas quanto à sua sublocalização celular, estrutura e ligante, assim como o são os mecanismos efetores acionados por eles e as vias de sinalização relacionadas (BRUBAKER et al., 2015 para uma revisão sobre PRRs).

A família de PRRs mais proeminente é a família de receptores semelhantes ao Toll (TLRs, *Toll-like receptors*), presentes em todos os animais, exceto planárias (Platyhelminthes) e anelídeos (Annelida) (KANZOK et al., 2004). Os TLRs são glicoproteínas que reconhecem padrões moleculares como ácidos nucleicos e lipídeos de bactérias,

fungos, protozoários e vírus (HASHIMOTO et al., 1988; LEMAITRE et al., 1996). São exclusivamente transmembranares, de superfície celular ou endossomais, e são expressos não apenas por células fagocitárias mas também por células epiteliais e fibroblastos nos mamíferos (CARIO et al., 2000).

Além dos TLRs, outras famílias de PRRs constituem a frente de ativação da resposta imunológica, como os receptores semelhantes à lectina tipo C (CLRs, *C-type lectin receptors*), os receptores semelhantes ao domínio de oligomerização ligante de nucleotídeo (NLRs, *nucleotide-binding oligomerization domain (NOD)-like receptors*), os receptores semelhantes à proteína induzida por ácido retinoico I (RLRs, *retinoic-acid-inducible gene I (RIG-I)-like receptors*) e os receptores *Scavenger*, para mencionar alguns (MANSUR; SMITH; FERGUSON, 2014; YUEN; BAYES; DEGNAN, 2013). Todos esses possuem correspondentes desde as esponjas, exceto os RLRs, que aparentemente surgiram logo antes do surgimento dos primeiros vertebrados, embora alguns grupos tenham sofrido perdas gênicas particulares, como é o caso dos nematódeos e artrópodes (BUCHMANN, 2014; FREEMAN et al., 1990; MUKHERJEE; KORITHOSKI; KOLACZKOWSKI, 2014; YUEN; BAYES; DEGNAN, 2013).

Os filos animais são separados por milhões de anos de evolução, logo, suas estratégias imunológicas variam conforme a história evolutiva particular de cada grupo. Sendo assim, cada filo carrega uma coleção de idiossincrasias, por exemplo, o número de receptores em cada família², as moléculas intermediárias das vias de sinalização relacionadas a cada receptor, o tipo de resposta efetora *et cetera*.

1.2.3 Sinalização celular e sistemas efetores imunobiológicos

Processos celulares são finamente coordenados por redes de reações moleculares. A orquestração das respostas imunológicas é realizada por centenas de fatores de regulação, citocinas e outras

2 O *Strongylocentrotus purpuratus*, o ouriço-do-mar, representante do filo Echinodermata, tem 222 genes de TLRs e o *Branchiostoma floridae*, o anfioxus, tem 71 cópias, enquanto os humanos temos 12 TLRs identificados (HUANG et al., 2008; RAST et al., 2006).

moléculas mediadoras, muitas das quais atuam via receptores celulares autólogos e parálogos que, por sua vez, acionam vias de sinalização apropriadas à resposta imunológica.

Os componentes fundamentais das redes de sinalização imunológicas que observamos nas espécies atuais consistem em domínios proteicos gerados precoce- e individualmente ao longo da evolução da vida. Esses módulos coalesceram nos Bilateria – antes da divergência entre Protostomia e Deuterostomia há cerca de 700 milhões de anos – a ponto de formar uma rede de sinalização funcional arquetípica a qual foi expandida e diversificada durante a evolução das linhagens animais (LIONGUE; SERTORI; WARD, 2016).

Geralmente se assume que nos vertebrados a rede de sinalização fundamental herdada do ancestral animal comum foi submetida a um aumento incrível de complexidade, embora se desconsidere que a maioria esmagadora da informação que temos sobre os sistemas imunobiológicos provenha dos estudos de, sobretudo, duas espécies de vertebrados, *Homo sapiens* e *Mus musculus*, enquanto continuamos ignorantes em relação à totalidade de estratégias adaptativas das demais espécies (DZIK, 2010). Todavia, eventos de duplicação genômica global nos vertebrados ocorreram pelo menos duas vezes (MEYER; SCHARTL, 1999). Algumas cópias, oriundas das expansões, foram mantidas nas espécies atuais e integram o sistema imunológico, por exemplo, os TLRs, as moléculas transdutoras de sinal e ativadoras de transcrição (STATs, *signal transduction and activators of transcription*), os Interferons (IFN) e seus receptores (LIONGUE; SERTORI; WARD, 2016).

Nos invertebrados, respostas celulares e humorais se combinam para conter infecções. Os mecanismos amplamente distribuídos nos filos são fagocitose, formação de cápsulas de melanina sobre agentes invasores, produção de peptídeos antimicrobianos (e.g. lisozimas, amoebaporinas), espécies reativas de oxigênio (ROS) e nitrogênio e coagulação da hemolinfa (LEIPPE et al., 1994; NAPPI; OTTAVIANI, 2000; SÖDERHÄLL; CERENIUS; JOHANSSON, 1994). A coagulação é um processo resultante da degranulação de hemócitos, semelhante ao que praticam neutrófilos nos vertebrados, que resulta na imobilização e morte de microrganismos invasores da hemolinfa (BRINKMANN et al., 2004; MUTA; IWANAGA, 1996). Apesar de tratarmos sobre a

imunidade de invertebrados de forma panorâmica, é importante ressaltar que existem particularidades em cada grupo e que generalizações certamente são equívocas.

O surgimento dos peixes mandibulados (Gnathostomata) representa um marco na evolução dos sistemas imunobiológicos, pois neles aparecem pela primeira vez na árvore da vida os MHCs, as células B e T e o mecanismo de geração de diversidade somática dependente dos genes de ativação de recombinação 1 e 2 (RAGs, *recombination activating gene*) que geram os receptores das células B e T e os anticorpos produzidos por essas (DAVIS et al., 1984; HOZUMI; TONEGAWA, 1976; MURAMATSU et al., 2000; OETTINGER et al., 1990; SCHLUTER et al., 1999). Um sistema semelhante a este, porém não-homólogo, existe nos peixes não-mandibulados (Agnatha), o que estabelece a existência de pelo menos três sistemas imunológicos adaptativos distintos nos seres vivos (PANCER et al., 2004).

Nos vertebrados, a percepção primária de padrões microbianos aciona redes de sinalização que incluem fatores de regulação de IFN (IRFs, *Interferon regulatory factors*) e vários fatores de transcrição [e.g. fator nuclear κ B (NF- κ B, *nuclear factor* κ B), proteína ativadora 1 (AP-1, *activator protein* 1)] que direcionam o perfil de expressão transcricional celular para uma resposta imunobiológica apropriada ao estímulo inicial, por meio da ativação da transcrição de citocinas pró-inflamatórias e IFNs os quais são secretados para o ambiente extracelular (MANSUR; SMITH; FERGUSON, 2014). A captação dessas moléculas pela própria célula e por suas vizinhas via receptores especializados aciona novas vias de sinalização que mobilizam mecanismos de contenção microbiana, entre eles, o aumento da expressão de moléculas de superfície (e.g. MHCs de classe I), a indução de genes/proteínas pró-apoptóticas (e.g. caspases), modulação da diferenciação celular e atividade citotóxica e muitos outros, os quais, enfim, resolvem a situação inflamatória que os ativou no primeiro momento (ORTALDO et al., 1983; GREINER et al., 1987).

1.2.4 A enzima cis-aconitato descarboxilase e seu produto, o ácido itacônico

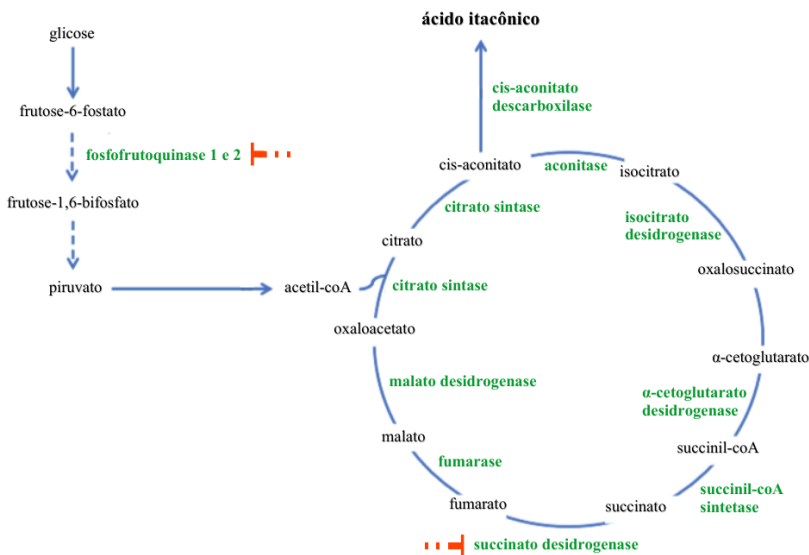
Entre as centenas de moléculas mobilizadas pelas células fagocitárias de vertebrados após o contato com sinais inflamatórios está o *Immune responsive gene 1* (IRG1), cuja expressão pode ser detectada cerca de 30 min³ após o contato com agentes infecciosos bacterianos, virais e parasíticos (BASLER et al., 2011; HALL et al., 2013; LI et al., 2006; MICHELUTTI et al., 2013; PREUSSE et al., 2011; SMITH et al., 2011; TALLAM et al., 2016).

Transcrito e traduzido o IRG1 é a enzima mitocondrial cis-aconitato descarboxilase (CAD). A CAD catalisa a conversão do metabólito cis-aconitato, derivado do ácido cítrico, em itaconato, também chamado ácido itacônico (AI), o qual é transportado para o citosol e, supostamente, externalizado, apesar de não se conhecer seu(s) transportador(es) (**Figura 1**) (BENTLEY; THIESSEN, 1957a; BENTLEY; THIESSEN, 1957b; DEGRANDI et al., 2009; MEISSNER et al., 2013). Ao longo dos últimos anos, os cientistas se interessaram em revelar o papel antimicrobiano e imunomodulatório envolvido nessa atividade enzimática (CORDES et al., 2016; MICHELUTTI et al., 2013).

O AI atua como inibidor da isocitrato liase, a enzima que aciona o ciclo do glioxilato, uma via alternativa de obtenção energética empregada durante condições de depleção nutricional, utilizada por arqueas, bactérias, fungos, plantas e protistas (CORDES; MICHELUTTI; HILLER, 2015; MCFADDEN; PUROHIT, 1977). O ciclo do glioxilato é uma estratégia adaptativa determinante para a sobrevivência de microrganismos patogênicos intracelulares que afetam vertebrados, como o *Legionella pneumophila*, *Mycobacterium tuberculosis* e *Salmonella enterica*, pois permite que essas bactérias utilizem compostos como ácidos graxos e colesterol ao invés de glicose, indisponível no ambiente intracelular (MICHELUTTI et al., 2013; NAUJOKS et al., 2016).

3 A expressão do IRG1 é detectada cerca de 40 min após estimulação pró-inflamatória de macrófagos humanos derivados de células progenitoras da medula óssea e 20 min depois em macrófagos murinos (TALLAM et al., 2016).

Figura 1 – Contexto bioquímico do ácido itacônico (AI). O substrato primário para a biossíntese do AI é a glicose. Ao ser obtida do meio extracelular é convertida a piruvato via glicólise e este adentra o ciclo do ácido cítrico como acetil coenzima A. O composto intermediário cis-aconitato, na presença da cis-aconitato descarboxilase, é convertido em AI. As enzimas envolvidas nos processos estão coloridas em verde e os metabólitos em preto. Os sinais em laranja indicam os dois pontos de bloqueio metabólico executado pelo AI sobre as enzimas fosfofrutoquinase 2 e succinato desidrogenase.



Fonte: modificada pela autora a partir de imagem publicada por KLEMENT; BÜCHS, 2013.

Adicionalmente, em peixes, o IRG1 exerce atividade antimicrobiana ao estimular a produção de ROS pelas mitocôndrias, atividade regulada pela via de sinalização Janus quinase/STAT (JAK/STAT, *Janus kinase*/STAT) e por glicocorticoides (HALL et al., 2013). Em mamíferos, a expressão de IRG1 é parcialmente dependente do fator de transcrição IRF1, embora ainda não se tenha maiores detalhes sobre a via (TALLAM et al., 2016).

No hospedeiro, o AI produzido durante a infecção interrompe a via glicolítica ao inibir a enzima fosfofrutoquinase 2, a qual realiza um dos primeiros passos do ciclo ao converter frutose-6-fosfato em frutose-2,6-bifosfato (SAKAI et al., 2004). A presença de frutose-2,6-bifosfato ativa a enzima fosfofrutoquinase 1 que fosforila o substrato frutose-6-fosfato em frutose-1,6-bifosfato. Em condições normais, esta daria continuidade ao ciclo (Figura 1). O bloqueio exercido pelo AI, semelhante ao efeito promovido sobre os microrganismos invasores, impede a síntese de ácidos graxos a partir de glicose (SAKAI et al., 2004).

O AI inibe ainda outra enzima-chave para o metabolismo do hospedeiro, a succinato desidrogenase (BOOTH et al., 1952). Esta enzima integra o ciclo do ácido cítrico ao converter succinato a fumarato. Esta inibição provoca o acúmulo de succinato nas mitocôndrias, fenótipo observado no contexto de respostas imunológicas inatas, o qual, por sua vez, estimula o eixo pró-inflamatório dependente da citocina IL-1 β , além de estimular a produção de ROS (TANNAHILL et al., 2013; LAMPROPOULOU et al., 2016). Como consequência, a respiração mitocondrial se torna comprometida.

Antes de os cientistas se debruçarem sobre a investigação do papel imunomodulatório e antimicrobiano do AI, outros já haviam se interessado por sua capacidade de polimerização. A presença de um grupo metileno na estrutura do AI lhe confere a possibilidade de autopolimerização, característica que passou a ser explorada industrialmente desde a metade do século passado como matéria-prima substituta para produtos químicos derivados do petróleo (EL-IMAM; DU, 2014; STEIGER et al., 2013). Entre as aplicações biotecnológicas do AI estão: tratamento de água, confecção de tintas, desodorantes, adesivos, resinas, plásticos, fibras de vidro e muitas outras.

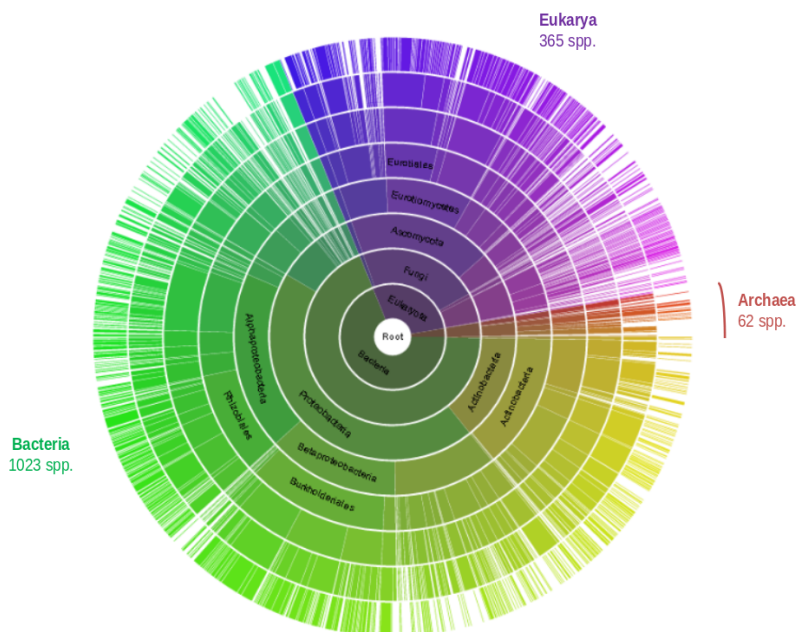
O AI utilizado na indústria é extraído de fungos. Fungos do gênero *Aspergillus* produzem até 80 g/l de AI e essa capacidade é suficiente e excedente à demanda comercial. A síntese de AI foi identificada pela primeira vez nesses fungos (KINOSHITA, 1931 *apud* CALAM; OXFORD; RAISTRICK, 1939; KLEMENT; BÜCHS, 2013). O interesse econômico motivou a pesquisa bioquímica a descrever as reações enzimáticas envolvidas na biossíntese de AI na intenção de manipulá-las a fim de incrementar sua taxa de produção (BONNARME

et al., 1995; YAHIRO *et al.*, 1995). As reações enzimáticas envolvidas na produção do AI em mamíferos revelaram-se serem as mesmas que as fúngicas (STRELKO, 2013).

1.3 PERGUNTA

Haja vista à semelhança bioquímica da síntese de AI e à distância evolutiva entre fungos e vertebrados de cerca de 1,3 bilhões de anos, nos perguntamos como se distribuiria a CAD no restante da árvore da vida (PARFREY *et al.*, 2011).

Figura 2 – Árvore filogenética da família de proteínas MmgE/PrpD indica lacunas na distribuição da CAD ao longo dos três domínios da vida.



Fonte: extraída da base de dados de domínios proteicos *online Pfam* <<https://goo.gl/vJwJ9t>>.

Preliminarmente, recorreremos à base de dados de domínios proteicos *online* Pfam (FINN et al., 2016). A CAD pertence à família de proteínas MmgE/PrpD, cujas integrantes, apesar de se estenderem por toda a árvore da vida, surpreendentemente, estão ausentes em grupos proeminentes, como o filo Arthropoda (**Figura 2**).

O que explica a distribuição observada?

Este trabalho é o estudo de caso de um gene, da proteína transcrita por ele, do produto enzimático desta e dos fenótipos que coletivamente constroem. Nós analisamos as informações disponíveis sobre o IRG1, CAD e AI em bancos de dados públicos e na literatura científica e, junto à análises bioinformáticas, investigamos os fatores que influenciaram sua evolução ao longo da árvore da vida.

1.4 OBJETIVOS

1.4.1 Objetivo geral

Investigar a trajetória evolutiva da CAD e sugerir causas para as múltiplas perdas gênicas observadas ao longo da árvore da vida.

1.4.2 Objetivos específicos

- a) Construir a árvore filogenética das espécies que possuem CAD a partir de sequências proteicas disponíveis *online*;
- b) Identificar, a partir da árvore construída, eventos evolutivos pontuais ao longos dos ramos;
- c) Buscar na literatura científica causas para os eventos evolutivos observados.

2 METODOLOGIA

2.1 ÁRVORES FILOGENÉTICAS

2.1.1 Obtenção e preparação das sequências

Entre as várias plataformas de registro de sequências proteicas disponíveis *online*, escolhemos a *evolutionary genealogy of genes: Non-supervised Orthologous Groups* (eggNOG) (<<http://eggnoг.embl.de>>) como fonte das sequências que utilizamos em nossas análises. A eggNOG processa genomas e proteomas provenientes de múltiplas bases de dados públicas [(Ensembl, UniProt Knowledgebase, Reference Sequence (RefSeq) e Joint Genome Institute (JGI) Genome Portal)] e os disponibiliza em uma coleção organizada de grupos ortólogos de proteínas e respectivas anotações funcionais (HUERTA-CEPAS et al., 2016).

A busca pelo elemento “IRG1” na página da eggNOG retorna a árvore de ortologia da família de proteínas MmgE/PrpD, disponível no endereço <<https://goo.gl/taFnnW>>. No total, 919 sequências provenientes de 538 espécies foram identificadas como ortólogas. Para a lista completa dos códigos de acesso, informações taxonômicas e demais informações sobre as 919 sequências, veja o **Apêndice A**. Todas as sequências foram descarregadas no mês de agosto de 2018 e a partir delas compomos nosso banco de dados para as análises filogenéticas subsequentes.

A preparação dos dados foi realizada no *shell*⁴ do sistema operacional Unix e a construção das árvores filogenéticas no *software* estatístico R (<<https://cran.r-project.org/>>). As linhas de comando determinantes para as análises filogenéticas estão registradas no **Apêndice B**.

2.1.2 Alinhamento e geração de árvores

Todas as sequências obtidas da eggNOG foram alinhadas com o programa Clustal Omega (SIEVERS et al., 2017). O Clustal Omega compara sequências com base em uma avaliação de distâncias gênicas.

⁴ Interface do sistema operacional UNIX que interpreta linhas de comando.

De forma breve, seu funcionamento se dá da seguinte forma: as N sequências de entrada são associadas por um processo de *embedding* (do inglês, “incorporação”) a vetores em um espaço de n dimensões, em que n é equivalente a $\log N$. O *embedding* é feito de tal maneira que o posicionamento dos vetores no espaço reflete, da melhor maneira possível, as relações entre as sequências (BLACKSHIELDS et al., 2010). Os vetores são então agrupados por meio dos métodos filogenéticos tradicionais k-means (MACQUEEN, 1967) e *Unweighted Pair-Group Method Using Arithmetic Average* (UPGMA) (SOKAL; MICHENER, 1958) em uma árvore-guia, cujas probabilidades são então calculadas com base em modelos ocultos de Markov (SÖDING, 2005).

A confiabilidade da árvore estimada pelo Clustal Omega foi quantificada com o pacote estatístico Pvcust (SUZUKI; SHIMODAIRA, 2006), implementado no ambiente R. O Pvcust calcula dois valores de probabilidade para cada agrupamento dos dados por meio de *bootstrapping* (FELSENSTEIN, 1985): a probabilidade de *bootstrap* (BP, *bootstrap probability*) e a probabilidade aproximadamente imparcial (AU, *approximately unbiased*). A partir desses dois testes, um valor entre 0 e 1 é atribuído a cada agrupamento hierárquico e este número é o valor de probabilidade que representa a possibilidade de o agrupamento ser verdadeiro.

Por fim, as árvores filogenéticas foram editadas para apresentação com o programa FigTree (versão 1.4.3) (<<https://goo.gl/pJPhTn>>). Devido a uma limitação do programa FigTree, empregado na edição das figuras, os ramos terminais das árvores não puderam ser destacados individualmente. Na versão atual do *software* é possível destacar apenas nós das árvores, que representam duas entidades taxonômicas ou mais. Por esse motivo e para fins ilustrativos, algumas espécies destacadas na árvores filogenéticas são meramente representativas de outras. As eventualidades em que essa improvisação foi necessária estão registradas em quadros indicados na legenda das respectivas figuras.

2.2 CLADOGRAMAS

Todos os cladogramas foram construídos no programa Mesquite (<<https://www.mesquiteproject.org>>) a partir de compilação teórica de filogenias extraídas da literatura científica (tópico 2.4) e não foram baseados em alinhamento de sequências.

2.3 HEATMAP

O *heatmap* das proteínas CAD e citramalil coenzima A liase foi construído na plataforma R com a função `heatmap.2` do pacote `Gplots` (<<https://goo.gl/3kPygg>>). As linhas de comando empregadas estão registradas no Apêndice B. A figura final foi manualmente reunida.

Os demais resultados cujas metodologias não foram detalhadas aqui foram obtidos com o auxílio de técnicas bioinformáticas básicas no *shell* do sistema operacional UNIX e na plataforma R.

2.4 NOMENCLATURA TAXONÔMICA

Ante à divergência das classificações taxonômicas das bases de dados e artigos científicos, principalmente, em relação aos grupos de procariotos e eucariotos unicelulares, optamos utilizar determinados sistemas de classificação para alguns grupos. As escolhas que fizemos estão justificadas a seguir:

a) Bacteria, Archaea e Eukarya: com a crescente expansão do inventário genômico publicado a megaclassificação dos organismos procarióticos ganhou várias versões. Optamos por seguir uma classificação tradicional, em que o domínio Eukarya é supostamente derivado de Archaea, embora as evidências recentes sugiram que Bacteria e Archaea tenham constituído os primeiros eucariotos, tornando-o um domínio intermediário aos outros dois (WILLIAMS et al., 2013). Os filós considerados para cada domínio são os sugeridos pela *List of Prokaryotic names with Standing in Nomenclature* (LPSN) (PARTE, 2014);

b) Eukarya: combinamos as classificações sugeridas por BURKI, 2014 e RUGGIERO et al., 2015, exceto em relação aos protozoários, discutidos no próximo tópico;

c) Protozoa: por não haver necessidade de aprofundamento, a classificação que assumimos para protozoários foi grosseira. Basicamente, reunimos todos os eucariotos não-animais e não-fungos sob o título Protozoa, ignorando a divisão de super grupos de Eukarya que os classifica em: Stramenopila, Alveolata e Rhizaria (SAR), Archaeplastida, Excavata, Amoebozoa, Opisthokonta e uma série de pequenos grupos menos resolvidos, como Haptophyta e Cryptomonada (BURKI, 2014; PARFREY et al., 2011). Os organismos abordados genericamente neste trabalho como “protistas” pertencem a três desses grandes grupos: SAR, Amoebozoa e Opisthokonta;

d) Chordata: tradicionalmente Chordata é considerado um filo, fragmentado em três subfilos: Vertebrata, Urochordata (Tunicata) e Cephalochordata. Com base em diferenças larvais fundamentais se sugeriu classificar Ambulacraria e Chordata como superfilos, logo, Echinodermata e Hemichordata se tornam filios de Ambulacraria e Cephalochordata, Urochordata e Vertebrata filios de Chordata (SATO, ROKHSAR e NISHIKAWA, 2014). Assumimos esta classificação neste trabalho;

e) Vertebrata: assumimos a classificação dos vertebrados proposta por DUNN et al., 2014.

2.5 LIMITAÇÕES DA PESQUISA

A geobiota, rica, diversa e repleta de seres e interações desconhecidas aos humanos, torna grandiosa a tarefa de investigar todos os seus elementos e sistemas. Cerca de 1,8 milhão de espécies foram identificadas até a data em que este manuscrito foi escrito. Dessas espécies, pouco menos de 15 mil tiveram seus genomas sequenciados e publicados (MUKHERJEE et al., 2016). A estimativa do número real de espécies de eucariotos é de 8,7 milhões (MORA et al., 2011). Existem previsões que variam entre 3 e 100 milhões de espécies, ou seja, de qualquer forma, os cientistas estão estudando muito menos de 1 % das espécies.

Além da limitação imposta pela informação registrada

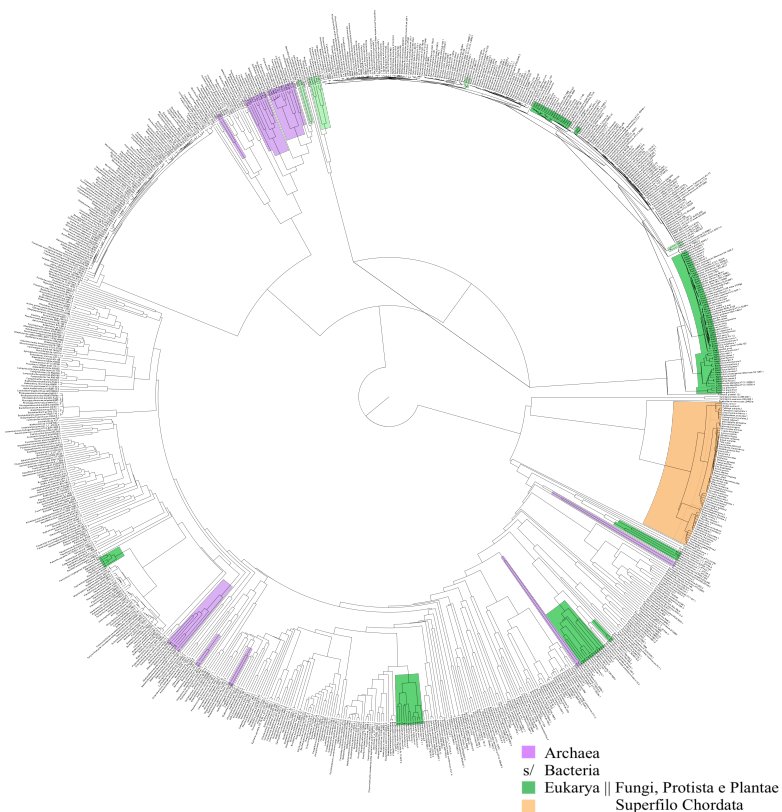
publicamente, contamos com a supremacia de detalhamentos de mamíferos e de algumas poucas espécies consideradas modelos de estudo, como *Drosophila* spp., *Caenorhabditis elegans* e *Danio rerio*. Essas proporções não são triviais ao pensarmos estudos de genética comparativa, pois, paradoxalmente, desconhecemos o que desconhecemos. É uma surpresa não encontrar exceções e particularidades biológicas ao investigar padrões fenotípicos ou genômicos, mas não temos como considerá-los com vistas à incompletude da informação biológica de que dispomos. Então, nesta seção de resultados, apesar de nos referirmos globalmente à todas as espécies do planeta, nossas conclusões são meras extrapolações.

O número de sequências que utilizamos nas análises moleculares é apenas uma amostragem das sequências que existem nas populações atuais. Além disso, os genomas dos indivíduos das espécies carregam uma coleção de polimorfismos, mantendo a variação intrínseca das espécies (FELIX, 2006). Além disso, o interesse monotônico nas espécies *Homo sapiens* e *Mus musculus* deixam as bases de dados relativamente pobres em recursos sobre as demais espécies. Enfim, as análises moleculares, por mais completas que pareçam, ignoram uma fatia importante de dados.

3 RESULTADOS E DISCUSSÃO

3.1 AS PROTEÍNAS DA FAMÍLIA MmgE/PrpD ESTÃO PRESENTES NOS TRÊS DOMÍNIOS DA VIDA

Figura 3 – A família de proteínas MmgE/PrpD exibe ampla distribuição nos três domínios da vida. Esta árvore filogenética foi inferida a partir do alinhamento de 919 sequências da família de proteínas MmgE/PrpD provenientes de 538 espécies, consideradas ortólogas pela base de dados eggNOG. O alinhamento foi realizado com o alinhador Clustal Omega e analisado estatisticamente com o pacote Pvclust. A ima- (continua na página 31)



Fonte: elaborada pela autora.

Quadro 1 – Substituições de espécies nas colorações da árvore filogenética da CAD nos três domínios da vida (Figura 3). “Espécies” são os indivíduos que de fato deveriam ser coloridos e “Representantes” são os que foram coloridos apesar de não pertencer aos devidos táxons..

Espécies				Representantes		
Espécie	Código de acesso	Grupo		Espécie	Código de acesso	Grupo
1	<i>Fusarium oxysporum</i> 1	FOXG_00484P0	Eukarya : Fungi	<i>Anaerotruncus colihominis</i> DSM 17241	ANACOL_04085	Bacteria
				<i>Clostridiales bacterium</i> 1_7_47FAA 7	CBFG_04389	Bacteria
2	<i>Natronomonas pharaonis</i> DSM 2160	NP4820A	Archaea	<i>Sphaerobacter thermophilus</i> DSM 20745 1	Sthe_1789	Bacteria
	<i>Haladaptatus paucihalophilus</i> DX253 2	ZOD2009_21077	Archaea	<i>Candidatus Koribacter versatilis</i> Ellin345	Acid345_2966	Bacteria
3	<i>Nectria haematococca</i> 2	NechaP89405	Eukarya : Fungi	<i>Clostridiales bacterium</i> 1_7_47FAA 2	MED121_13270	Bacteria
				<i>Clostridiales bacterium</i> 1_7_47FAA 5	PSPTO_3057	Bacteria
4	<i>Emericella nidulans</i> 3	CADANIAP00008256	Eukarya : Fungi	<i>Salinispora arenicola</i> CNS-205 1	Sare_1027	Bacteria
				<i>Salinispora tropica</i> CNB-440 1	Strop_4385	Bacteria
5	<i>Halorubrum lacusprofundi</i> ATCC 49239 1	Hlac_1088	Archaea	<i>Halorubrum lacusprofundi</i> ATCC 49239 1	Hlac_1088	Archaea
				<i>Ramlibacter tataouinensis</i> TTB310 1	Rta_00810	Bacteria
6	<i>Halorubrum lacusprofundi</i> ATCC 49239 2	Hlac_1089	Archaea	<i>Methylobacterium</i> sp. 4-46 4	M446_1932	Bacteria
				<i>Azorhizobium caulinodans</i> ORS 571	AZC_2114	Bacteria
7	<i>Tetrapisipora phaffii</i> CBS 4417	XP_003688641.1	Eukarya : Fungi	<i>Pseudogulbenkiana ferrooxidans</i> 2002	FuraDRAFT_1817	Bacteria
				<i>Verminephrobacter eiseniae</i> EF01-2.2	Veis_0651	Bacteria
8	<i>Perkinsus marinus</i> ATCC 50983	XP_002772992.1	Eukarya	<i>Candidatus Nitrospira defluvii</i>	NIDE3125	Bacteria
	<i>Monosiga brevicollis</i> MX1	XP_001743258.1	Eukarya : Choanoflagellida	<i>Chthoniobacter flavus</i> Ellin-428	CfE428DRAFT_5043	Bacteria

Fonte: elaborado pela autora.

gem final foi editada no *software* FigTree (versão 1.4.3). As espécies de arqueas estão representadas em roxo, as bactérias foram mantidas sem destaque e os eucariotos foram subdivididos em: laranja para os cordados, verde escuro para fungos e verde claro para os demais eucariotos (plantas, protistas e coanoflagelados). O **Quadro 1** relaciona algumas espécies que foram coloridas mas que não são integrantes dos respectivos táxons. Vide Metodologia (tópico 2.1.2) para mais detalhes. A árvore estendida contendo os valores de suporte estatístico de cada nó e a tabela contendo os mesmos estão disponíveis como **Apêndices C e D**, respectivamente.

Na análise de relações evolutivas, comumente se assume o princípio da parcimônia, em que uma hipótese coerente de árvore é aquela que demonstra o relacionamento entre as unidades biológicas que compara com o menor número de passos evolutivos possível. Ou seja, a ocorrência de uma única mudança evolutiva é considerada mais provável, ou mais parcimoniosa, que duas mudanças ou mais. Ao deixar-se guiar por esse princípio, as evidências que coletamos dos seres vivos atuais nos indica, por exemplo, que toda a vida na Terra deriva de um ancestral celular comum (DARWIN, 1859; WOESE; FOX, 1977). Todos os seres vivos compartilham características essenciais que sustentam seus processos bioquímicos fundamentais, por exemplo, um mecanismo de hereditariedade baseado em ácidos nucleicos. Tamanha é a complexidade de alguns sistemas que a probabilidade de os seres vivos terem surgido de forma independentemente múltiplas vezes é considerada improvável.

Embora o princípio da parcimônia simplifique a tomada de decisões durante a análise de hipóteses evolutivas, a parcimônia, por vezes travestida de economia energética, não é uma propriedade intrínseca dos seres, mas apenas um conceito filosófico aplicado a um método de previsão (SOBER, 1994). Apesar de ter sido postulado que as hipóteses mais simples têm maior probabilidade de estarem corretas, é importante considerar que subestimamos a incidência de mudanças evolutivas nas populações. Por exemplo, ao compararmos duas sequências nucleicas distantes entre si por certo tempo evolutivo, podemos identificar quais sítios foram alterados e qual a direção da mudança, entretanto, ignoramos quaisquer outras alterações intermediárias transcorridas durante o meio tempo.

De qualquer forma, considerando o pressuposto da descendência comum, infere-se a partir da Figura 3 que o ancestral comum à toda a vida no planeta possuía uma cópia ancestral da CAD. Todas as sequências semelhantes à CAD presentes nos organismos atuais são, portanto, descendentes dessa cópia primária. Mais além, podemos sugerir que as espécies que atualmente não possuem a CAD, certamente a perderam em algum momento. Esta conclusão é mais parcimoniosa que supormos o surgimento de sequências e funções extremamente parecidas ou idênticas duas ou mais vezes ao longo da evolução das espécies, apesar de não podermos descartar essa possibilidade.

Um aspecto notável da Figura 3 é a fragmentação dos níveis taxonômicos maiores. O alinhamento das sequências da família MmgE/PrpD não reproduziu o agrupamento esperado para os domínios, reinos e filos conforme a classificação taxonômica consenso.

Não é incoerente que uma árvore de sequências não tenha a mesma topologia que a árvore de especiação dos organismos considerados. Algumas propriedades dos seres e da técnica de construção da árvore nos ajudam a entender o porquê dessa fragmentação.

Primeiro, uma árvore de especiação leva em consideração dados genéticos, morfológicos e paleontológicos e os aproxima para supor a história evolutiva de uma espécie ou nível taxonômico maior. Já uma árvore de sequências se limita ao alinhamento de ácidos nucleicos ou proteínas e o valida por meio de métodos matemáticos. Por partirem de tipos diferentes de informação, seus resultados são essencialmente distintos.

Segundo, a existência dos genes é marcada por um processo perene de alterações ao longo das gerações. Quanto maior o tempo evolutivo que se estende desde o surgimento de um gene ou espécie, maior o número de substituições, deleções, inserções de origem interna ou externa⁵ e recombinações que se acumulam em sua estrutura (KIMURA, 1968; ZUCKERKANDL; PAULING, 1965). Esses processos distanciam sequências e espécies inicialmente próximas, sem que necessariamente resultem em alterações de função ou eventos de especiação. Por exemplo, na Figura 3, as CADs das espécies de

5 Transferência gênica horizontal, processo ubíquo entre seres unicelulares (DOOLITTLE, 1999).

Aspergillus, em 1,3 bilhões de anos de evolução dos fungos, se distanciaram a ponto de formar quatro clados distantes entre si, quando comparadas com as sequências das demais espécies (PARFREY et al., 2011).

Terceiro, sequências gênicas não se alteram apenas via mutações e recombinações randômicas ao longo das gerações. As populações passam por remodelamentos gênicos em resposta direta ao ambiente e às pressões seletivas pelas quais os indivíduos passam (EHRlich; RAVEN, 1969; FELIX, 2006). Os organismos inclusos na Figura 3 são fenotipicamente muito distintos, assim como seus habitats e modos de vida. Considere por um momento os estilos de vida da *Cavia porcellus*, a capivara, um roedor que chega a 100 kg que habita regiões próximas a rios, lagos e pântanos, e do *Toxoplasma gondii*, protozoário intracelular obrigatório de metacelulares. Cada ser vivo é regido por redes biológicas complexas e intimamente interconectadas que respondem dinamicamente ao ambiente e à sua própria variação interna e a CAD, nesse contexto, é apenas um elemento dentre os milhares que as compõem. As mudanças pelas quais as espécies passaram ao longo de suas trajetórias evolutivas pressionaram pontos diferentes das redes e o resultado de todos os movimentos de acomodação compreende o genoma atual de cada espécie.

3.2 AS PROTEÍNAS DA FAMÍLIA MmgE/PrpD CONSTITUEM DUAS SUBFAMÍLIAS DISTINTAS

Em arqueas não há trabalhos específicos sobre a atividade enzimática ou qualquer outro detalhamento sobre a CAD e por isso, infelizmente, não as abordaremos no texto devido à falta de informação.

Nas bactérias em que se têm estudo, por exemplo em *Salmonella typhimurium* e *Escherichia coli*, as proteínas da família MmgE/PrpD atuam como 2-metilcitrato desidratases. Nessas espécies, o gene que as codifica é integrante do operon prpRBCDE, envolvido no catabolismo do propionato, subproduto da degradação de compostos carbonados e tóxico para bactérias (BLANK; GREEN; GUEST, 2002, BERG; FILATOVA; IVANOVSKY, 2002, BROCK et al., 2002). A atividade catalítica da PrpD sobre o cis-acnitato, substrato da CAD de mamíferos

e fungos, é insignificante (HORSWILL; ESCALANTE-SEMENERA, 1999).

O sistema de classificação *Protein ANalysis THrough Evolutionary Relationships* (PANTHER) indica que todas as “CADs” bacterianas, assim como as de alguns fungos (e.g. *Emericella nidulans*, *Phaeosphaeria nodorum*) e eucariotos basais (e.g. *Dictyostelium discoideum*, *Paramecium tetraurelia*), são classificadas sob a mesma subfamília (<<https://goo.gl/Cmy6tX>>). De acordo com a descrição do método de classificação do PANTHER, famílias são grupos de proteínas relacionadas evolutivamente e subfamílias são proteínas relacionadas que desempenham a mesma função. Portanto, as proteínas da família MmgE/PrpD se distinguem em dois módulos distintos, em conformidade com as funções que executam.

A partir de agora, para fins de distinção, nos referiremos às proteínas integrantes da subfamília das 2-metilcitrato desidratases como “CADs ancestrais” (CADas).

Entre os eucariotos considerados na árvore filogenética da CAD, apenas duas espécies de plantas foram reconhecidas pela eggNOG como portadoras da CAD: *Ostreococcus tauri* e *Ostreococcus lucimarinus*. Essas algas unicelulares são os menores eucariotos de vida livre conhecidos, medindo entre 0.8 e 1.2 µm, e participam da constituição do picofitoplâncton (DEMIR-HILTON et al., 2011; GUIRY; GUIRY, 2018). Não se têm estudos específicos sobre a CAD dessas espécies, logo, não sabemos que função desempenham nesse grupo.

A representação mais numerosa entre os eucariotos na Figura 3 se dá por parte dos fungos, com 66 espécies, seguidos pelos metazoários, com 49 espécies. Há registro de enzimas de ambas subfamílias nos fungos. As espécies consideradas pelo PANTHER como portadoras da CADa pertencem todos ao filo Ascomycota, filo mais recente dentre os fungos, com origem há aproximadamente 600 milhões de anos (PARFREY et al., 2011).

Os fungos do gênero *Aspergillus* (Filo Ascomycota) foram os primeiros espécimes em que a produção de AI foi identificada (CALAM; OXFORD; RAISTRICK, 1939). Não se investigou até agora a manutenção da produção de AI nos fungos a partir de uma perspectiva evolutiva, entretanto, sabe-se que o AI inibe o seu crescimento

(JAKLITSCH; KUBICEK; SCRUTTON, 1991; STEIGER et al., 2013; YAHIRO et al., 1995).

Sete protistas foram incluídos na filogenia, dos quais três estão incluídos na classificação do PANTHER como portadores de CADas. São eles: *Dictyostelium discoideum*, *Dictyostelium purpureum* e *Paramecium tetraurelia*. Além destes, consta também como integrante da subfamília ancestral o coanoflagelado *Monosiga brevicollis*.

Os demais protistas inclusos na filogenia que, entretanto, não foram mencionados no PANTHER, pertencem ao supergrupo Alveolata e são eles: *Neospora caninum*, *Perkinsus marinus*, *Tetrahymena thermophila* e *Toxoplasma gondii*. Não temos informações específicas sobre a função da CAD nesses organismos.

Podemos concluir a partir desta seção que os seres unicelulares simples, ancestrais das espécies atuais, detinham uma cópia primária da CAD, a qual foi herdada pelo ancestral comum do supergrupo Opisthokonta, existente por volta de 1 bilhão de anos atrás e que deu origem aos fungos e aos animais (PARFREY et al., 2011). Nos fungos, a CAD tomou dois caminhos: na maior parte das espécies manteve a função ancestral e nas demais se especializou em converter cis-aconitato a itaconato. Nos animais, a CAD assumiu exclusivamente esta função.

Quadro 2 – Sequências resultantes do alinhamento da CAD humana com a ferramenta BLAST. As informações contidas neste quadro foram extraídas diretamente da página de resultados do alinhamento na BLAST e da página de cada sequência na base de dados Genbank. Id %: identidade; Pred: sequência predita; S/rev: notação sem revisão; SA: sítios ativos.

Filo	Espécie	Código de acesso	Nome da proteína	Score	E-value	Id %	Pred	S/rev	SA	Observações
Arthropoda	<i>Drosophila obscura</i>	XP_022212151.1	uncharacterized protein LOC111067402	55.8	5e-06	21	x			0/8
	<i>Drosophila elegans</i>	XP_017129775.1	PREDICTED: 2-methylcitrate dehydratase 1-like	55.8	5e-06	21				0/8
	<i>Drosophila fuscipennis</i>	XP_017045444.1	PREDICTED: 2-methylcitrate dehydratase 1-like	55.5	7e-06	21				0/8
	<i>Drosophila obscura</i>	XP_022234573.1	uncharacterized protein LOC111082655	55.1	8e-06	22				0/8
	<i>Drosophila kikkawai</i>	XP_017029022.1	PREDICTED: 2-methylcitrate dehydratase 1-like	53.5	3e-06	21				0/8
	<i>Drosophila bipectinata</i>	XP_017087944.1	PREDICTED: 2-methylcitrate dehydratase 1-like	53.5	3e-06	21				0/8
	<i>Drosophila serrata</i>	XP_020806223.1	uncharacterized protein LOC110182517	52.4	6e-06	22				0/8
	<i>Drosophila eugracilis</i>	XP_017062732.1	PREDICTED: 2-methylcitrate dehydratase 2-like	51.2	1e-04	22				0/8
	<i>Drosophila fuscipennis</i>	XP_017045452.1	PREDICTED: 2-methylcitrate dehydratase 1-like	45.1	0.005	36				0/8
	<i>Stylophora pistillata</i>	PFX34234.1	Protein FAM208A	34.3	2.0	31				0/8
Cnidaria	<i>Stylophora pistillata</i>	XP_022787984.1	uncharacterized protein LOC111327928	34.3	2.21	31				0/8
	<i>Stylophora pistillata</i>	XP_022788243.1	putative beta-lactamase-like 1	32.3	5.9	47				0/8
	<i>Nematostella vectensis</i>	XP_001640824.1	predicted protein	28.9	9.5	30				0/8 Fragmento parcial (66 aa)
	<i>Stylophora pistillata</i>	PFX27052.1	Beta-lactamase-like protein 3	32.0	9.8	48				0/8
Cephalochordata	<i>Branchiostoma floridae</i>	XP_002597324.1	hypothetical protein BRAFLDRAFT_203633	438	2e-150	50		x		7/8
	<i>Branchiostoma belcheri</i>	XP_019618481.1	PREDICTED: cis-aconitate decarboxylase-like	423	2e-144	48		x		7/8
Mollusca	<i>Biomphalaria glabrata</i>	XP_013076153.1	PREDICTED: cis-aconitate decarboxylase-like	434	4e-147	47				7/8
	<i>Lotia gigantea</i>	XP_009054769.1	hypothetical protein LOTGIDRAFT_232322	431	5e-146	47		x		7/8
	<i>Mizuhopecten yessoensis</i>	XP_021359565.1	cis-aconitate decarboxylase-like isoform X2	428	4e-145	46	x	x		7/8
	<i>Mizuhopecten yessoensis</i>	XP_021359566.1	cis-aconitate decarboxylase-like isoform X3	427	5e-145	46	x	x		7/8
	<i>Mizuhopecten yessoensis</i>	XP_021359564.1	cis-aconitate decarboxylase-like isoform X1	428	5e-145	46	x	x		7/8
	<i>Pomacea canaliculata</i>	PVD23427.1	hypothetical protein C0Q70_16696	413	4e-139	48				7/8
	<i>Crassostrea virginica</i>	XP_022312773.1	cis-aconitate decarboxylase-like	410	2e-138	44	x			7/8
	<i>Crassostrea virginica</i>	XP_022319601.1	cis-aconitate decarboxylase-like	410	4e-138	44	x			7/8
	<i>Octopus bimaculoides</i>	XP_014779084.1	PREDICTED: cis-aconitate decarboxylase-like	394	8e-132	44				7/8
	<i>Crassostrea gigas</i>	XP_011438844.1	PREDICTED: cis-aconitate decarboxylase	390	2e-130	43				7/8
	<i>Crassostrea gigas</i>	EKC41910.1	hypothetical protein CGI_10013925	389	5e-130	43				7/8
	<i>Ostrea edulis</i>	AFA34413.1	Irg1 immunoresponsive protein 1	53.9	3e-08	38				0/8 Fragmento parcial (73 aa)
	<i>Opisthorchis viverrini</i>	XP_009167745.1	hypothetical protein T265_13562	31.6	8.5	38		x		0/8
	<i>Schistosoma haematobium</i>	XP_012796880.1	Neuroendocrine convertase 2	31.6	8.6	32		x		0/8

Fonte: elaborado pela autora.

3.3 MOLUSCOS, CEFALOCORDADOS E VERTEBRADOS SÃO OS ÚNICOS FILOS DE METAZOA EM QUE CONSTA A PRESENÇA DA CIS-ACONITATO DESCARBOXILASE

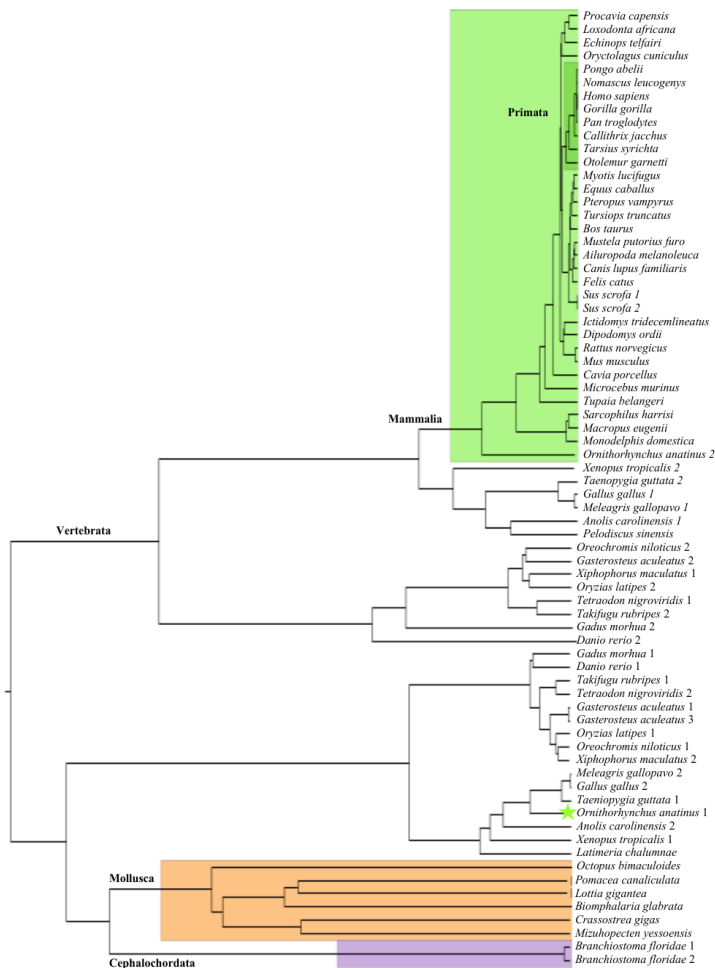
Para corroborar a amplitude da distribuição da CAD nos metazoários sugerida pelos dados da eggNOG, realizamos uma busca alternativa. Nós alinhamos a CAD humana a todas as sequências de todos os filos de Metazoa disponíveis na página da *Basic Local Alignment Search Tool* (BLAST) (<<https://goo.gl/jiHnpV>>) e dentre os 34 filos pareados, os que resultaram alinhamentos significativos foram: Arthropoda, Cephalochordata, Cnidaria, Mollusca, Platyhelminthes e Vertebrata. De acordo com a eggNOG, apenas os filos Cephalochordata e Vertebrata possuem proteínas ortólogas à CAD. A fim de resolver essa divergência, analisamos individualmente as sequências encontradas com a nova busca.

Os alinhamentos foram analisados relativamente quanto à sua qualidade por meio dos valores de *score*, *e-value* e identidade extraídos da BLAST e à presença dos sítios ativos da CAD, preditos para a sua atividade enzimática nos humanos (MICHELUTTI et al., 2013). O resumo da análise pode ser consultado no **Quadro 2**.

De acordo com nossa avaliação, apenas os filos Cephalochordata, Mollusca e Vertebrata possuem a enzima CAD. Todas as sequências de Mollusca, exceto o fragmento parcial de *Ostrea edulis*, apresentaram valores significativos para a qualidade do alinhamento (em relação à CAD humana) e todas possuem sete dos oito sítios ativos preditos para a CAD (**Apêndice E**).

A partir de nossa análise, constatamos que as proteínas do filo Mollusca são putativamente ortólogas às CADs. Supomos que a não inclusão do filo no grupo de ortólogos da eggNOG tenha ocorrido devido à alguma restrição no algoritmo para filtragem de sequências da base de dados. A árvore da família de proteínas MmgE/PrpD construída pela base de dados Pfam, por sua vez, inclui o filo Mollusca, o que corrobora nossa análise (<<https://goo.gl/fhHEAj>>). A Pfam inclui também em sua filogenia o filo Nematoda, entretanto, a partir de nossa análise justificamos a não inclusão deste no grupo de ortólogas da CAD.

Figura 4 – Relações filogenéticas da CAD nos metazoários sugere a ocorrência de 16 eventos de duplicação. Esta árvore foi inferida a partir do alinhamento de 72 seqüências da CAD, presentes em 34 espécies de metazoários. O alinhamento foi realizado com o alinhador Clustal Omega e analisado com o pacote Pvcust. A imagem final (continua na próxima página)...



Fonte: elaborada pela autora.

foi editada no *software* FigTree (versão 1.4.3). O filo Cephalochordata está colorido em roxo, o filo Mollusca em laranja e a classe Mammalia em verde, com pequeno destaque para o clado dos primatas. Uma pequena estrela verde indica uma espécie de mamífero (*Ornithorhynchus anatinus* 1) deslocada do grande grupo. A árvore estendida contendo os valores de suporte estatístico, assim como a tabela com os mesmos estão disponíveis nos **Apêndices F e G**, respectivamente.

3.4 AS CIS-ACONITATO DESCARBOXILASES PARÁLOGAS ENCONTRADAS NOS METAZOÁRIOS APARENTEMENTE MANTÊM A MESMA FUNÇÃO

Ao analisarmos a distribuição da CAD nos filios de metazoários, notamos a ocorrência de duplicações em 16 espécies de vertebrados (**Figura 4**). Em nove casos, uma das cópias sofreu translocação cromossômica.

Genes relacionados entre si por eventos de especiação são ditos ortólogos e genes relacionados entre si por um evento de duplicação são frequentemente chamados de parálogos (FITCH, 1970). Teoricamente, é comum considerar genes parálogos como as principais causas de emergência de novas funções gênicas. Imaginou-se, durante a expansão teórica da genética molecular na década de 70, que imediatamente após duplicação, uma das cópias gênicas sofre afrouxamento das pressões seletivas sobre sua função e, como consequência, passa a evoluir mais rapidamente. Esta situação culminaria, então, em perda de função ou neofuncionalização da cópia e, neste caso, sua taxa evolutiva diminuiria novamente (OHNO, 1970).

Há registro na literatura científica de casos em que ambas cópias passam por redução das pressões seletivas e, como consequência do aumento de suas taxas evolutivas, tornam-se subunidades com função complementar e particionada, processo particularmente chamado de subfuncionalização (LYNCH, FORCE, 2000).

A análise comparativa das espécies *Homo sapiens* (ser humano), *Mus musculus* (camundongo), *Gallus gallus* (galinha), *Caenorhabditis elegans* (nematódeo), *Drosophila melanogaster* (mosca-da-fruta), *Arabidopsis thaliana* (planta herbácea), *Oryza sativa* (arroz) e *Saccharomyces cerevisiae* (levedura) permitiu a estimativa de uma taxa geral de surgimento de duplicações em que se espera que 50 % dos

genes em um genoma passe por um evento de duplicação pelo menos uma vez em um intervalo de 35-350 milhões de anos. Se estiver correta, essa estimativa pode indicar que a maioria dos genes duplicados aparentemente é neutralizada, ao invés de preservada via neo- ou subfuncionalização (LYNCH, CONERY, 2000). Dessa forma, a meia-vida média de um gene duplicado recai em aproximadamente 4 milhões de anos.

Nós analisamos cada caso de duplicação da CAD para ter maior garantia da existência de ambos genes. Supomos que pelo menos 9 dos 16 casos de paralogia registrados pela eggNOG são verídicos, pois as posições cromossômicas informadas no perfil de cada cópia é distinta. Não podemos afirmar que as demais são falsas, porém, não sabemos se as entradas das sequências são simplesmente notações alternativas da mesma proteína.

A análise individualizada das sequências duplicadas nos revela que todas possuem os sítios ativos preditos para a função enzimática da cis-aconitato descarboxilase e que a maioria mantém entre 50 e 60 % de identidade entre si (resultados não mostrados).

Curiosamente, as sequências parálogas de peixes, anfíbios, lagartos e aves se subdividiram em dois clados distintos. Talvez aqui estejamos observando o processo evolutivo sugerido canonicamente para sequências parálogas, em que uma das cópias sofre relaxamento seletivo e tende a alterar-se mais rapidamente que a outra.

De forma não surpreendente, o ornitorrinco se mostrou uma exceção dentre os mamíferos. Enquanto uma das cópias de sua CAD se agrupou da forma esperada junto aos mamíferos marsupiais, a outra cópia se agrupou a três aves e ao *Anolis carolinensis*, o lagarto. A anatomia e fisiologia do ornitorrinco, assim como o seu genoma, reflete o mosaico de características derivadas de seus ancestrais mamíferos e reptilianos (WARREN et al., 2008). No caso dessa espécie, talvez estejamos observando um processo evolutivo diferencial e paralelo das CADs, alimentado por pressões seletivas relacionadas a componentes regulatórios ou hábitos de ambos grupos ancestrais. A combinação inusitada de dois modos de vida característicos de grupos distintos no ornitorrinco possibilitou o que parece ser algo semelhante a uma subfuncionalização, em que cada cópia tornou-se semelhante à CAD de

cada grupo ancestral, a cópia dois aos mamíferos e a cópia um aos répteis. Porém, não temos mais evidências além dessas.

Levando em consideração a taxa de duplicação gênica estimada para os seres vivos e a noção teórica de relógio molecular, supomos que os parálogos da CAD que observamos atualmente não acumularam alterações nucleotídicas o bastante para que houvesse ganho ou perda de função e a manutenção dos sítios ativos em todas as sequências nos indica fortemente que a função enzimática permanece semelhante (DOS REIS et al., 2015; ZUCKERKANDL; PAULING, 1965).

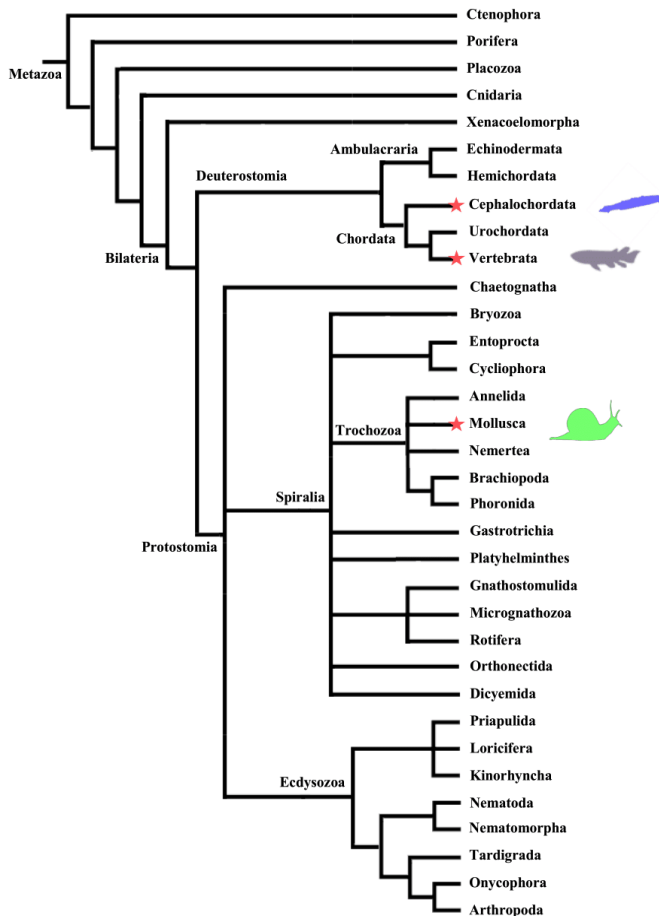
Há evidências (e controvérsias) de que pelo menos duas duplicações genômicas globais ocorreram nos vertebrados nos períodos pré-cambriano e devoniano, há aproximadamente 600 e 400 milhões de anos (MEYER; SCHARTL, 1999). Se a inferência de meia vida média de 4 milhões de anos para os genes estiver correta, é possível que os vertebrados ancestrais tenham passado por eventos de duplicação seguidos de neutralização das cópias e os parálogos que vemos atualmente seriam as únicas remanescentes.

Relações de ortologia e paralogia se somam às evidências para a predição de função proteica, entretanto, contrário ao que frequentemente se considera teoricamente, para muitas famílias de proteínas a maior parte dos parálogos mantém funções semelhantes por alguns milhões de anos para então tornar-se neutra, e é isso o que aparentemente observamos na história evolutiva da CAD nos animais (Figura 4) (LYNCH, CONERY, 2000; PEARSON, 2013). A exposição das CADs dos metazoários a um teste de detecção de seleção positiva direcional poderia fornecer algum *insight* sobre seus processos evolutivos recentes.

3.5 O GENE CODIFICADOR DA CIS-ACONITATO DESCARBOXILASE FOI PERDIDO MÚLTIPLAS VEZES AO LONGO DA EVOLUÇÃO

Até agora vimos que o ancestral comum aos três domínios da vida na Terra carregava um(a) gene/proteína que com o passar de alguns milhões de anos se especializou em duas funções distintas. Nas bactérias, nos protistas e em alguns fungos assumiu a função principal de degradação do propionato, um composto tóxico derivado do proces-

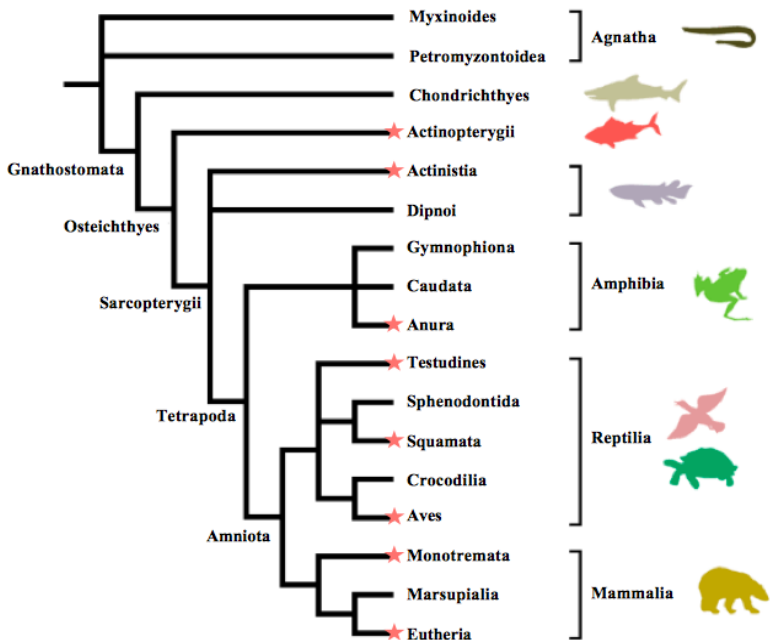
Figura 5A – O gene codificador da CAD foi perdido múltiplas vezes ao longo da história evolutiva dos animais. Este cladograma foi construído manualmente, independente de seqüências nucleicas ou aminoacidicas, a partir da classificação filogenética sugerida por DUNN et al., 2014. As estrelas vermelhas indicam a presença da CAD em espécies do grupo assinalado.



Fonte: elaborada pela autora.

samento de compostos carbonados (BLANK; GREEN; GUEST, 2002, BROCK et al., 2002). Nos metazoários e demais fungos, passou a atuar como enzima produtora de AI, o qual, nos mamíferos, foi identificado como sendo um metabólito antimicrobiano e imunomodulatório produzido por macrófagos ativados (CALAM; OXFORD; RAISTRICK, 1939; STRELKO et al., 2011; MICHELUTTI et al., 2013).

Figura 5B – O gene codificador da CAD foi perdido múltiplas vezes ao longo da história evolutiva dos vertebrados. As estrelas vermelhas indicam a presença da CAD em espécies do grupo assinalado. Este cladograma foi desenhado manualmente, independente de sequências nucleicas ou aminoacídicas.



Fonte: modificada pela autora a partir de imagem disponibilizada pela *University College London* em <<https://goo.gl/Tir7GQ>>.

A partir de agora nos debruçaremos exclusivamente sobre os aspectos envolvidos na evolução da forma recente da CAD. Surpreendentemente, se compararmos os resultados obtidos com a Figura 3 a uma árvore de especiação compreendendo as espécies atuais, percebemos que há muitas lacunas na distribuição da CAD nas espécies (**Figura 5A e Figura 5B**).

Não podemos estimar com segurança a porcentagem de espécies que perderam a CAD em relação a todas as espécies do planeta devido à falta de dados sobre espécies não-modelos (vide Metodologia, tópico 2.5, Limitações da pesquisa), porém, se compararmos sua distribuição com o número de espécies cujos genomas foram sequenciados e publicados (aproximadamente 15 mil), percebemos que as espécies positivas para a CAD compreendem menos de 4 % (MAKHERJEE et al., 2016).

Entende-se perda aqui como quaisquer modificações gênicas que descaracterizaram o gene codificador da CAD e o tornam identificável no genoma, embora mudanças mais tímidas possam inviabilizar o fenótipo produzido pela CAD, como o fechamento epigenético da cromatina, a alteração do promotor ou a perda de fatores de transcrição. Apesar de não podermos garantir a produção de AI por todas as versões da CAD que detectamos, tomaremos esse fato como pressuposto.

O que, então, poderia ter ocasionado as múltiplas perdas do gene codificador da CAD ao longo da árvore da vida?

3.6 O ÁCIDO ITACÔNICO TEM EFEITO TÓXICO

In vitro, doses suprafisiológicas de AI controlam populações de bactérias como *Legionella pneumonophila*, *Mycobacterium tuberculosis* e *Salmonella enterica*, protagonistas de infecções intracelulares, e *in vivo*, mais de 75 % dos camundongos deficientes para IRG1, quando desafiados com *Mycobacterium*, sucumbem à infecção em até 30 dias (MICHELUTTI et al., 2013; NAIR et al., 2018; NAUJOKS et al., 2016). Nestes animais, a mortalidade está associada a maior infecção, inflamação e patologia.

O efeito antimicrobiano exercido pelo AI é resultado de sua atividade inibitória sobre o ciclo do glioxilato, bacteriano, requisito para a permanência no ambiente intracelular do hospedeiro (DUNN;

RAMÍREZ-TRUJILLO; HERNÁNDEZ-LUCAS, 2009; MCFADDEN; PUROHIT, 1977; MUÑOZ-ELÍAS, 2006).

3.6.1 A citramalil-coenzima A liase é provavelmente ubíqua no processo de degradação do ácido itacônico

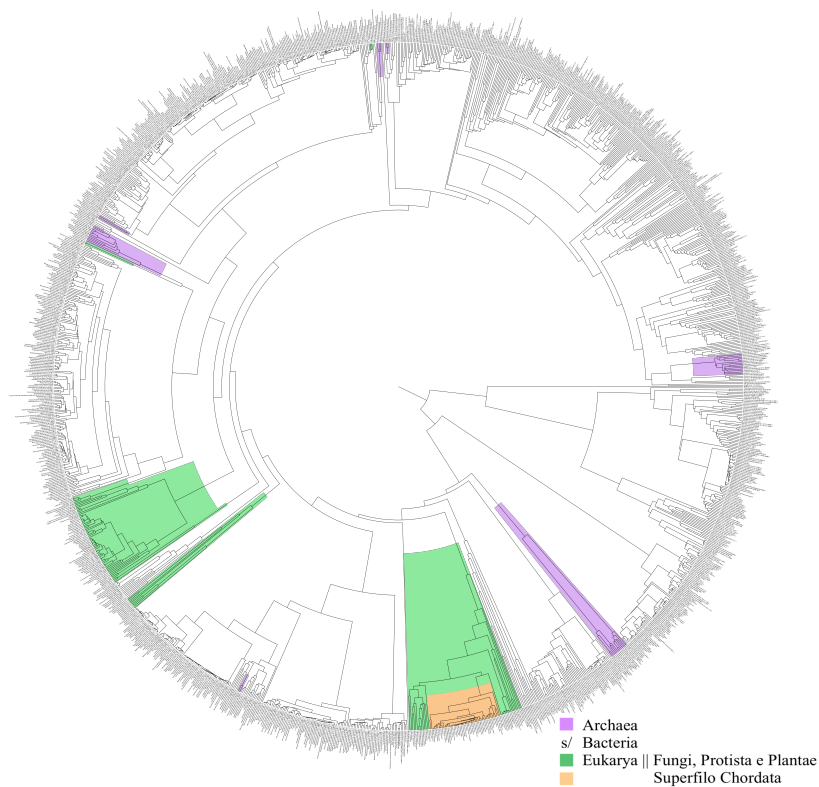
Algumas espécies de bactérias, como *Yersinia pestis* e *Pseudomonas aeruginosa*, possuem a habilidade de degradar o AI por meio da ação combinada de três enzimas: itaconato coenzima A (coA) transferase, itaconil coA hidratase e citramalil coA liase (Ccl) (SASIKARAN et al, 2014). Várias outras espécies de bactérias possuem genes homólogos a essas proteínas, por exemplo as espécies *Legionella longbeachae*, *Mycoplasma synoviae* e *Salmonella enterica*, o que pode indicar a presença do mesmo fenótipo (CORDES; MICHELUTTI; HILLER, 2015).

Da mesma forma, as enzimas mitocondriais de células hepáticas de mamíferos também são capazes de degradar o AI (ADLER; WANG; LARDY, 1957; WANG; ADLER; LARDY, 1961). Entretanto, os dois primeiros passos da via são executados, respectivamente, pelas enzimas não-específicas a essa função: succinil coA sintetase e metilglutaconase (ADLER; WANG; LARDY, 1957; WANG; ADLER; LARDY, 1961).

A Ccl humana é homóloga à bacteriana e é expressa na matriz mitocondrial (CALVO et al., 2016; DIMROTH; EGGERER, 1975). As Ccls pertencem à família de proteínas HpcH/Hpal e estão presentes nos três domínios da vida, assim como a CAD (**Figura 6**). A família de subdivide em quatro subfamílias, entretanto, não exploraremos suas particularidades.

Semelhante à árvore filogenética da CAD, algumas espécies destacadas na Figura 6 não pertencem ao grupo taxonômico a que foi atribuída, mas representam outras espécies que não puderam ser coloridas devido à uma falha no *software* FigTree utilizado para edição das árvores. As substituições para eventual consulta estão registradas no **Quadro 3**.

Figura 6 – Relações filogenéticas da Ccl indicam ampla distribuição nos três domínios da vida. Esta árvore foi inferida a partir do alinhamento de 1471 sequências da família de proteínas HpcH/HpaI provenientes de 842 espécies, consideradas ortólogas pela base de dados eggNOG. O alinhamento foi realizado com o alinhador Clustal Omega e analisado estatisticamente com o pacote Pvcust. A imagem final foi editada no *software* FigTree (versão 1.4.3). As espécies de arqueas estão representadas em roxo, as bactérias foram mantidas sem destaque e os eucariotos foram subdivididos em: laranja para os cordados, verde escuro para fungos e verde claro para os demais eucariotos (plantas, protistas e coanoflagelados). A árvore estendida contendo os valores de suporte estatístico de cada nó e o quadro contendo os mesmos estão disponíveis, respectivamente, nos **Apêndices H e I**.



Fonte: elaborada pela autora.

Quadro 3 – Espécies substituídas nas colorações da árvore filogenética da Ccl (Figura 6). As espécies pertencentes aos grupos taxonômicos destacados estão indicadas apenas como “Espécies” e ao lado estão as espécies que foram elegidas representantes das mesmas e, logo, destacadas na figura.

Espécies				Representantes		
	Espécie	Código de acesso	Grupo	Espécie	Código de acesso	Grupo
1	<i>Aciduliprofundum boonei</i> T469	Aboo_0969	Archaea	Candidatus <i>Koribacter versatilis</i> Ellin345	Acid345_2963	Bacteria
				<i>Aciduliprofundum boonei</i> T469	Aboo_0969	Archaea
2	<i>Naegleria gruberi</i>	XP_002678699.1	Eukarya	<i>Desulfotobacterium hafniense</i> Y51.2	DSY3039	Bacteria
				<i>Desulfosporosinus orientis</i> DSM 785.1	Desor_1258	Bacteria
3	<i>Haladaptatus paucinalophilus</i> DX253.1	ZOD2009_16041	Archaea	<i>Corynebacterium ulcerans</i> BR-AD22	CULC22_00706	Bacteria
				<i>Corynebacterium pseudotuberculosis</i> I19.1	Cpl19_0650	Bacteria

Fonte: elaborado pela autora.

Até a presente data, apenas um trabalho abordou mecanisticamente a função *in vivo* da Ccl em mamíferos. A partir dele temos que a Ccl realiza a função altamente específica de conversão do composto citramalil-coA em acetil coA e piruvato, embora trabalhos mais antigos do mesmo grupo tenham demonstrado outras atividades enzimáticas de menor coeficiente cinético (SHEN et al., 2017). Possivelmente essas atividades alternativas integram o metabolismo de outras espécies.

3.6.2 Todos os metazoários portadores de cis-aconitato descarboxilase possuem também citramalil-coenzima A liases

A partir dos recortes da literatura mencionados nesta seção, bolamos a seguinte hipótese: o AI é um composto tóxico, não apenas para os microorganismos que oportunamente se instalam em células fagocitárias animais, mas também para o hospedeiro produtor de AI. Dessa forma, a presença da CAD em um sistema biológico requereria um mecanismo de controle negativo, por exemplo, uma via de degradação, tal como foi descrita na literatura científica para bactérias e mamíferos (ADLER; WANG; LARDY, 1957; WANG; ADLER; LARDY, 1961; SASIKARAN et al., 2014).

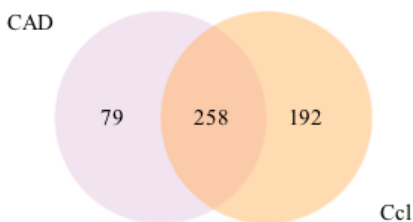
Para investigar essa hipótese, analisamos o genoma das espécies animais em busca da coexistência da CAD e da Ccl, a enzima responsável pelo último passo da via de degradação do AI. Para tanto, buscamos a página da família de proteínas HpcH/HpaI, a qual a Ccl pertence, na base de dados eggNOG e descarregamos a lista de espécies que contêm a proteína (disponível em <<https://goo.gl/x5Rrmc>>). Reduzimos a lista a apenas gêneros e fizemos o mesmo para a CAD. No total, comparamos 450 gêneros portadores da Ccl e 337 gêneros portadores da CAD cujo resultado pode ser visto nas **Figuras 7A e 7B**.

A partir da Figura 7B temos que em todos os animais com dados disponíveis, as enzimas CAD e Ccl coexistem. Assumindo a funcionalidade das enzimas em todos os casos, nossos resultados sugerem que todas as espécies produtoras de AI são capazes de degradá-lo.

Como explorado no tópico 3.2, CADs e CADas constituem subfamílias distintas e desempenham funções não relacionadas. A

CADa não tem como produto o AI, logo, a essencialidade de degradação desse composto não se faz necessária. A diversidade de estados de presença/ausência das enzimas observada nos seres vivos não-animais (Figura 7B) é uma evidência da distinção da atividade desempenhada pelas proteínas das duas subfamílias.

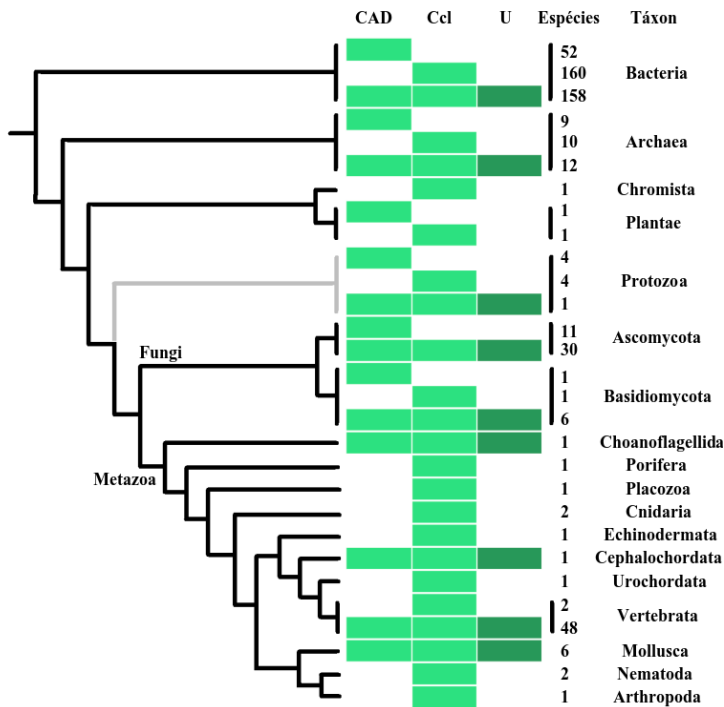
Figura 7A – As proteínas CAD e Ccl coexistem em 258 gêneros nos três domínios da vida. O círculo roxo, à direita, representa os gêneros que possuem a CAD em seus genomas e o círculo laranja, à esquerda, representa os gêneros que possuem Ccl. A intersecção entre os dois inclui os gêneros em que ambas proteínas coexistem. Esta figura foi gerada no ambiente R com o pacote venn.



Fonte: elaborada pela autora.

As demais espécies, portadoras da CAD e que, porém, não possuem Ccl, são todas protozoários unicelulares. *Neospora caninum* e *Toxoplasma gondii* são parasitos intracelulares obrigatórios de mamíferos, *Perkinsus marinus* é um parasito intracelular facultativo de moluscos e *Tetrahymena thermophila* é um ciliado de vida livre dulcícola. Supomos que estes, caso suas proteínas integrem a subfamília de MmgE/PrpD derivada, produtoras de AI, possuam outras maneiras de controle da toxicidade do AI, por exemplo, a possibilidade de secreção, como observado em fungos, sobretudo por serem unicelulares (STEIGER et al., 2013).

Figura 7B – Todos os animais portadores de CAD possuem Ccl. Esta figura representa os estados em que (co)existem as proteínas CAD e Ccl em 529 gêneros dos três domínios. Os retângulos verdes claros indicam a presença das respectivas enzimas e os retângulos verdes escuros indicam a coexistência (U) de ambas no mesmo organismo. Alguns grupos taxonômicos são heterogêneos em relação aos estados de (co)existência das proteínas. São eles: Bacteria, Archaea, Plantae, Protozoa, Ascomycota, Basidiomycota e Vertebrata. Por exemplo, em Bacteria, 52 gêneros possuem apenas CAD, 161 possuem apenas Ccl e 158 possuem ambas enzimas. Os táxons foram indicados conforme relevância para a nossa pergunta e por isso não correspondem ao mesmo nível taxonômico. Esta figura foi construída sequencialmente: o cladograma foi desenhado com o *software* Mesquite baseado em filogenia consenso da literatura científica e não em alinhamento de seqüências, o *heatmap* foi gerado no ambiente R com a função heatmap.2 do pacote Gplots e a figura final foi montada manualmente.



Fonte: elaborada pela autora.

3.7 A AUSÊNCIA DE UM MECANISMO DE CONTROLE NEGATIVO DA PRODUÇÃO DE ÁCIDO ITACÔNICO TORNA A PRESENÇA DA CIS-ACONITATO DESCARBOXILASE INCOMPATÍVEL COM A VIDA DOS METAZOÁRIOS

Por serem enzimas amplamente distribuídas ao longo dos três domínios da vida supomos que o ancestral comum aos animais possuía ambas CAD e Ccl. Temos, nos vertebrados, apenas duas condições: espécies que as mantiveram e espécies que perderam CAD mas mantiveram Ccl. Os cenários hipotéticos de coevolução das enzimas a partir do ancestral CAD⁺Ccl⁺ ao longo do tempo são:

a) Alguns organismos as herdaram de seus ancestrais e as mantêm até hoje, por exemplo, os humanos;

b) Em alguns, a perda da Ccl foi seguida pela perda da CAD, ou alternativamente, a perda da CAD foi seguida pela perda não relacionada da Ccl. Um desses casos certamente aconteceu nos vertebrados que não apresentam nenhuma das enzimas, como crocodilianos e salamandras;

c) Em dois casos particulares a CAD foi perdida e a Ccl se manteve, nas espécies vertebradas *Dasypus novemcinctus*, o tatu, e *Macaca mulatta*, o macaco-rhesus.

Em qualquer dos três cenários imaginários, a existência da CAD é obrigatoriamente acompanhada da Ccl.

O genoma dos répteis é caracterizado pela presença de muitos macro- e microcromossomos. Em aves, grupo derivado dos répteis, foi identificado que os microcromossomos são genicamente densos e apresentam taxas altas de recombinação, cinco vezes maiores que os observados em cromossomos de mamíferos, o que sugere que os micromossomos são fatores importantes de geração de variação e, portanto, a perda de ambas CAD e Ccl em vários répteis (Crocodylia, Sphenodontia, Caudata e Gymnophiona) (Figura 5B) pode ser parcialmente explicada pelas particularidades genômicas do grupo (ORGAN; MORENO; EDWARDS, 2008).

Em relação ao tatu e ao macaco-rhesus, a perda da CAD aparentemente não implicou em comprometimento adaptativo das espécies em relação ao módulo evolutivo que estamos analisando, pois a dependência entre as enzimas é unidirecional. Não encontramos na

literatura nenhuma informação relevante sobre particularidades imunológicas em ambas espécies.

Nossa análise comparativa demonstrou a correspondência da distribuição das enzimas CAD e Ccl nos genomas dos metazoários. Nós sugerimos, a partir desta observação e do conjunto de informações coletadas neste manuscrito, que essa correlação representa um fenótipo adaptativo importante para a maior parte dos grupos de vertebrados. Na ausência da Ccl os indivíduos CAD⁺ enfrentaram condições de vulnerabilidade evolutiva por estarem suscetíveis tanto à toxicidade do AI quanto a infecções por agentes patogênicos, sobretudo, bacterianos, e é por esse motivo que não existem vertebrados CAD⁺Ccl⁻.

Todavia, para ser confirmada, nossa hipótese requer investigação empírica. Uma forma de abordar a essencialidade da Ccl na manutenção da CAD nas espécies é desafiar *in vitro* células de espécies negativas para Ccl com AI e observar a mortalidade celular, a qual seria o indicador direto da tolerância das células para com o metabólito.

Um trabalho recente demonstrou uma relação sólida entre as enzimas CAD e Ccl nos humanos. Humanos possuem Ccl, exceto cerca de 2,7 % da população mundial que apresenta uma versão inativa da enzima devido à inserção de um códon de parada precoce nos ancestrais (SHEN et al., 2017). Nenhum sintoma da inativação da Ccl é perceptível a não ser níveis reduzidos de B12 circulante no sangue. Na presença de AI extracelular, células fagocitárias humanas o capturam e o convertem intracelularmente a itaconil-coA. Ao interagir com a coenzima B12, o itaconil-coA a converte a cob(II)alamina, a qual é suscetível à inativação por oxidação (PADOVANI; BANERJEE, 2006; SHEN et al., 2017). A B12 é cofatora da enzima metilmalonil-coA mutase (MUT), uma enzima da matriz mitocondrial que catalisa a conversão de metilmalonil-coA a succinil-coA e que, dessa forma, participa da reciclagem de cadeias carbônicas (SHEN et al., 2017). Na ausência da Ccl, o itaconil-coA acumula-se, inativa a B12 e, no processo, inibe a MUT. O grupo concluiu que a função enzimática da Ccl é necessária para a manutenção da função da B12 mitocondrial por meio da “detoxificação” de itaconil-coA (SHEN et al., 2017).

As cobalaminas, os metabólitos da família da coenzima B12, são componentes essenciais para a sobrevivência, reprodução e bem-estar de todos os vertebrados e da grande maioria dos demais seres vivos. Os

únicos seres produtores de cobalaminas são procariotos (bactérias e algas) e todos os demais suprem suas necessidades fisiológicas de B12 ao extraírem-na do ambiente. Por exemplo, peixes que vivem em ambiente selvagem extraem uma boa dose de B12 ao se alimentarem de algas (BRAEKKAN, 1958). Os animais piscívoros, por sua vez, como mamíferos e outros peixes, obtêm as cobalaminas via ingestão desses predadores primários. Além da dieta, acredita-se que as cobalaminas são obtidas frequentemente a partir da microbiota associada ao indivíduo (DANCHIN; BRAHAM, 2017).

A dependência em relação a essa família de metabólitos se confirma por meio da observação de animais com dietas controladas. Galinhas e pombas alimentadas apenas com sementes sofrem de anemia causada pela deficiência de B12 (CASTLE, 1985). Primatas depletados igualmente desenvolvem anemia, assim como paralisia, apatia, problemas de visão, degeneração do tecido nervoso e, eventualmente, morte (AGAMANOLIS et al., 1976; CHESTER et al., 1980). *Caenorhabditis elegans* deficientes para B12 têm retardo no crescimento, tornam-se inférteis e falecem mais cedo que o esperado (BITO; WATANABE, 2016). Um hábito comum desenvolvido por muitos animais nutricionalmente deficientes é a coprofagia, por meio da qual se obtém um estoque farto de nutrientes, microrganismos e, certamente, B12 (ROSENBERG; ZILBER-ROSENBERG, 2016; SOAVE; BRAND, 1991).

Temos, a partir das informações reunidas, três exemplos de grupos taxonômicos distantes que sofrem com a deficiência das cobalaminas. O AI, ao reduzir os níveis de B12 circulante agride veementemente as atividades fisiológicas dos metazoários e, por isso, o controle negativo sobre o AI certamente é submetido a mais um ponto de pressão seletiva. A correlação que observamos entre as enzimas CAD e Ccl neste trabalho, portanto, ganhou um reforço.

3.8 A EVOLUÇÃO DA CIS-ACONITATO DESCARBOXILASE NOS VERTEBRADOS É MULTIFATORIAL

Nossa hipótese abarca um cenário possível para a distribuição da CAD observada nos seres vivos. Não atribuímos à nossa hipótese exclusividade causal para a manutenção da CAD nos genomas das espécies. A resposta à seleção de qualquer traço específico pode ser

compreendida apenas no contexto de toda a história de vida dos organismos alvos junto às características do arcabouço gênico que os compõe (SACCHERI; HANSKI, 2006). As células/espécies são redes de elementos intimamente relacionados e os coeficientes de seleção podem variar em resposta a qualquer componentes endógeno ou exógeno, inclusive a seleção de outro traço (SACCHERI; HANSKI, 2006).

Teoricamente é possível supor que qualquer variação fenotípica altera os demais traços do indivíduo em diferentes intensidades, e que qualquer traço é afetado por múltiplos fatores (KACSER; BURNS, 1981; WAGNER; PAVLICEV; CHEVERUD, 2007). Por exemplo, a ativação transcricional de um gene afeta os genes vizinhos, com funções não necessariamente relacionadas, em até 100 kb de distância (GHANBARIAN; HURST, 2015). O conjunto de fenótipos de um indivíduo estaria, portanto, mesmo que minimamente, sob efeito pleiotrópico generalizado, o que torna ingênua a atribuição de causas únicas aos fenômenos biológicos, facilmente observada nos títulos muitas publicações científicas.

A seguir exploraremos brevemente dois outros possíveis pontos de pressão seletiva sobre a trajetória evolutiva da CAD.

3.8.1 A ausência do *Immune responsive gene 1* durante o processo de implantação embrionário em camundongas implica em taxas abortivas de 50-100 %

Em camundongas grávidas, o processo de implantação dos embriões na parede do útero ocorre precisamente quatro dias após a fertilização dos óvulos (PARIA; HUET-HUDSON; DEY, 1993). Durante o período que transcorre entre esses dois eventos, o útero da mãe se torna receptivo à nidação após tratamento com progesterona durante 48 h seguida de um pico de estrogênio (PSYCHOYOS, 1976; PSYCHOYOS, 1986).

Um trabalho demonstrou que as células epiteliais do lúmen do útero passam a expressar IRG1 no segundo dia após a fertilização e que essa expressão se torna acentuada (20 vezes maior que no segundo dia) no quarto dia, coincidindo com a janela de tempo esperada para a nidação (CHEON et al., 2003). A expressão de IRG1 nessas células é

dependente de progesterona e na sua ausência, o processo de implantação embrionária é prejudicado. As camundongas sem IRG1 durante os primeiros dias após fertilização sofrem 50-100 % de redução do número de embriões implantados (CHEON et al., 2003).

As altas taxas abortivas na ausência de IRG1 sugerem uma forte seleção sobre os mamíferos placentários durante o início da gestação, o que seria suficiente para explicar a manutenção desse gene nos genomas desses animais.

3.8.2 Coevolução de microrganismos intracelulares e hospedeiros pode ter pressionado a manutenção da cis-aconitato descarboxilase em moluscos e vertebrados

Interações entre microrganismos patogênicos e seus hospedeiros se destacam como pressões evolutivas importantes sobre ambas espécies envolvidas. Os primeiros alteram de forma drástica a dinâmica bioquímica dos últimos ao interferir em funções fisiológicas básicas e impactam diretamente a sobrevivência das populações (FUMAGALLI et al., 2011).

Microrganismos infectantes intracelulares ocorrem em uma variedade de hospedeiros vertebrados e não-vertebrados e caracterizam-se por alterar o microambiente celular, tornando-o compatível com sua sobrevivência (DIACOVICH et al., 2007; FRANCO; SHUMAN; CHARPENTIER, 2009). Entre os microrganismos com essa tendência ressaltamos a relevância da bactéria *Salmonella enterica*, do protista alveolado *Perkinsus marinus* e do complexo de bactérias *Mycobacterium tuberculosis* (MTBC, *Mycobacterium tuberculosis complex*).

Salmonella infecta muitas espécies de vertebrados, incluindo répteis, aves e mamíferos, e insetos (EVANGELOPOULOU et al., 2013). *Perkinsus marinus* infecta uma variedade de moluscos e o MTBC tem uma amplitude de hospedeiros mamíferos ampla, incluindo bovinos, pinipédios (focas, leões-marinhos, lobos-marinhos e morsas) e seres humanos (SMITH et al., 2009). Destacamos esses microrganismos por terem regimes de infecção obrigatoriamente intracelulares e, enquanto nesse contexto, estão certamente suscetíveis à ação do AI produzido pelas células fagocitárias em que se alojam.

O efeito antimicrobiano do AI foi demonstrado sobre populações de *Mycobacterium* e *Salmonella in vitro* e *in vivo* (para a micobactéria), indicando participação parcial do metabólito na contenção da infecção (MICHELUTTI et al., 2013, NAIR et al., 2018). Assim como os vertebrados, os moluscos possuem CAD e, plausivelmente, produzem AI em resposta à infecções embora não se tenha evidências experimentais.

Sugerimos, de forma parcialmente especulativa, que os filos Vertebrata e Mollusca tenham passado por afunilamentos populacionais no passado durante a coevolução com os microrganismos mencionados, resultando em uma população que responde às infecções de forma mais tolerante que a ancestral (ou seja, há controle da infecção sem que haja morte do hospedeiro) (FUMAGALLI et al., 2011; GAGNEAUX, 2012). Dessa forma, ambas espécies (microrganismo e hospedeiro) sobrevivem. Devido ao papel imunológico desempenhado pelo AI sobre agentes patogênicos intracelulares, as espécies infectadas tiveram maior probabilidade de sobreviver na presença da CAD, talvez, impondo sobre a enzima certa pressão seletiva positiva.

3.9 CONCLUSÕES

Todos os seres vivos são laboratórios de experiências do desenvolvimento biológico e da seleção natural. Novidades fenotípicas complexas podem surgir em uma simples mudança de ambiente. Portanto, a biologia evolutiva inevitavelmente trata de causas proximais assim como de processos seletivos a fim de encontrar explicações satisfatórias para os processos de mudança das espécies (WEST-EBERHARD, 2003. p 197).

Neste trabalho tombamos nossa atenção sobre um módulo fenotípico: a enzima CAD e seu produto, o AI. A partir de informações disponibilizadas publicamente investigamos as causas da distribuição da CAD em todas as espécies que tivemos acesso. Nos deparamos com uma dicotomia na família de proteínas a que a CAD pertence, que as dividiu em dois grandes grupos com funcionalidades distintas. Os procariotos empregam uma versão da CAD que as auxilia na detoxificação de produtos fisiológicos. Nos demais seres vivos, apesar

de a especificidade da função enzimática não ser a mesma, o resultado final é igualmente a detoxificação de produtos fisiológicos.

Nos vertebrados, adicionalmente ao seu papel metabólico, a CAD passou a integrar o conjunto de moléculas relacionadas ao sistema imunológico.

Os peixes mandibulados são os seres de origem evolutiva mais antiga a carregarem o conjunto gênico determinante para o desenvolvimento da imunidade dos vertebrados. Devido ao desconhecimento dos cientistas sobre as formas de vida intermediárias entre agnatos e gnatostomados, o surgimento “repentino” de muitas estruturas imunológicas fundamentais há aproximadamente 500 milhões de anos parece, à nossa visão turva, como um *big bang* molecular (SCHLUTER et al., 1999). A recombinação somática que constroi os receptores de células B e T, os anticorpos, os MHCs e os IFNs são exemplos dessa explosão de diversidade (FLAJNIK; KASAHARA, 2001; KRAUSE; PESTKA, 2005; MARCHALONIS et al., 1998; RAST et al., 1997).

O IRG1, gene que codifica a CAD nos vertebrados, tem pelo menos um fator de transcrição estimulado por IFN, o IRF1. Supomos que os movimentos de acomodação das redes moleculares após o *big bang* imunológico tornaram a CAD uma enzima relacionada às funções imunológicas desses animais (CROMBACH; HOGEWEG, 2008).

Entretanto, nem todas as espécies atuais mantiveram o IRG1 em seus genomas. Buscar respostas do porquê foi o principal escopo deste trabalho. Identificamos uma forte correlação entre a perda da CAD e a ausência de uma via de degradação do AI. O efeito tóxico do AI sobre o organismo produtor requer um mecanismo de controle e as espécies que porventura o perderam (ou não o desenvolveram), sofreram a perda da(o) CAD/IRG1.

Adicionalmente, identificamos na literatura pontos de pressão seletiva adicionais sobre a evolução da CAD. O IRG1 desempenha um papel essencial, porém desconhecido, durante a implantação embrionária em camundongos e a coevolução patógeno-hospedeiro pode ter contribuído para a manutenção da CAD nas espécies de moluscos e vertebrados.

Trilhões de interações, em todos os níveis de organização dos seres, ocorrem a cada instante. Não conseguimos enquanto humanos

enxergar e compreender simultaneamente todas as nuances da vida. Rotineiramente, reduzimos metadados a duas dimensões e expandimos observações pontuais a modelos gerais. Igualmente humanos, seguimos essa mesma trajetória e das poucas evidências construimos nossa hipótese, certamente incompleta mas cautelosamente pensada.

Nota da autora: este trabalho foi realizado sem provocar o sofrimento ou morte de qualquer ser vivo. Da Declaração Universal dos Direitos dos Animais de 1978 proclamada pela UNESCO, “Art. 1º - Todos os animais nascem iguais perante a vida e têm os mesmos direitos à existência”.

REFERÊNCIAS

ADLER, J., WANG, S.F., LARDY, H.A. The metabolism of itaconic acid by liver mitochondria. **The Journal of Biological Chemistry**, v. 229, n. 2, p. 865-879, 1957.

AGAMANOLIS, D.P. et al. Neuropathy of experimental vitamin B12 deficiency in monkeys. **Neurology**, v. 26, p. 905-14, 1976.

ALBERTS, B. et al. **Biologia Molecular da Célula**. 5ª ed. Porto Alegre: Artmed, 2010. p. 2.

ALLEN, P.G., DAWIDOWICZ, E.A. Phagocytosis in *Acanthamoeba*: I. A mannose receptor is responsible for the binding and phagocytosis of yeast. **Journal of Cell Physiology**, v. 145, n. 3, p. 508-13, 1990.

ARNDT, N.T.; NISBET, E.G. Processes on the young Earth and the habitats of early life. **Annual Review of Earth and Planetary Sciences**, v. 40, p. 521-549, 2012.

BAASKE, P. et al. Extreme accumulation of nucleotides in simulated hydrothermal pore systems. **Proceeding of the National Academy of Sciences**, v. 104, n. 22, p. 9346-51, 2007.

BARRANGOU, R. et al. CRISPR provides acquired resistance against viruses in prokaryotes. **Science**, v. 315, n. 5819, p. 1709-12, 2007.

BASLER, T. et al. *Mycobacterium paratuberculosis*, *Mycobacterium smegmatis*, and lipopolysaccharide induce different transcriptional and post-transcriptional regulation of the IRG1 gene in murine macrophages. **Journal of Leukocyte Biology**, v. 79, p. 628-38, 2006.

BENGTSON, S. Origins and early evolution of predation. **Paleontological Society Papers**, v. 8, p. 289-317, 2000.

BENTLEY, R., THIESSEN, C.P. Bisynthesis of itaconic acid in *Aspergillus terreus*. I. Tracer studies with C14-labeled substrates. **Journal of Biological Chemistry**, v. 226, p. 673-78, 1957a.

BENTLEY, R., THIESSEN, C.P. Bisynthesis of itaconic acid in *Aspergillus terreus*. III. The properties and reaction mechanism of cis-aconitic acid decarboxylase. **Journal of Biological Chemistry**, v. 226, p. 703-20, 1957b.

BERG, I.A., FILATOVA, L.V., IVANOVSKY, R.N. Inhibition of acetate and propionate assimilation by itaconate via propionyl-CoA carboxylase in isocitrate lyase-negative purple bacterium *Rhodospirillum rubrum*. **Federation of European Microbiological Society Microbioly Letters**, v. 216, n. 1, p. 49-54, 2002.

BININDA-ELMONDSs, O.R.P. et al. The delayed rise of present-day mammals. **Nature**, v. 446, p. 507-12, 2007.

BITO, T., WATANAKE, F. Biochemistry, function, and deficiency of vitamin B12 in *Caenorhabditis elegans*. **Experimental Biology and Medicine**, v. 241, p. 1663-8, 2016.

BLACKSHIELDS, G. et al. Sequence embedding for fast construction of guide trees for multiple sequence alignment. **Algorithms for Molecular Biology**, v. 5, n. 21, 2010.

BLANK, L., GREEN, J., GUEST, J.R. AcnC of *Escherichia coli* is a 2-methylcitrate dehydratase (PrpD) that can use citrate and isocitrate as substrates. **Microbiology**, v. 148, p. 133-46, 2002.

BONNARME, P. et al. Itaconate biosynthesis in *Aspergillus terreus*. **Journal of Bacteriology**, v. 177, n. 12, p. 3573-8, 1995.

BOOTH, A.N. et al. The inhibitory effects of itaconic acid in vitro and in vivo. **Journal of Bacteriological Chemistry**, v. 195, n. 2, p. 697-702, 1952.

BRAEKKAN, O.R. Vitamin B12 in marine fish. **Nature**, v. 182, p. 1386, 1958.

BRINKMANN, V. et al. Neutrophil extracellular traps kill bacteria. **Science**, v. 303, n. 5663, p. 1532-5, 2004.

BROCK, M. et al. Oxidation of propionate to pyruvate in *Escherichia coli*: involvement of methylcitrate dehydratase and aconitase. **European Journal of Biochemistry**, v. 269, p. 6184-9, 2002.

BROWN, D.; GODDARD, N.L. Exponential DNA replication by laminar convection. **Physical Reviews Letters**, v. 91, n. 15, p. 158103-1-158103-4, 2003.

BRUBAKER, S.W. et al. Innate immune pattern recognition: a cell biological perspective. **Annual Review of Immunology**, v. 33, p. 257-90, 2015.

BRUSCA, R.C., BRUSCA, G.J. **Invertebrates**. 2nd. edition. Sunderland: Sinauer Associates, 2003. p. 48.

BUCHMANN, K. Evolution of innate immunity: clues from invertebrates via fish to mammals. **Frontiers in Immunology**, v. 5, n. 459, p. 1-8, 2014.

BURKI, F. The eukaryotic tree of life from a global phylogenomic perspective. **Cold Spring Harbor Perspectives in Biology**, v. 6, n. 5, 2014.

CALAM, C.T., OXFORD, A.E., RAISTRICK, H. Studies in the biochemistry of microorganisms: itaconic acid, a metabolic product of a strain of *Aspergillus terreus* Thom. **Biochemical Journal**, v. 33, n. 9, p. 1488-95, 1939.

CALVO, S.E., CLAUSER, K.R., MOOTHA, V.K. MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins. **Nucleic Acids Research**, v. 44, n. D1, D1251–D1257, 2016.

CARIO, E. et al. Lipopolysaccharide activates distinct signaling pathways in intestinal epithelial cell lines expressing Toll-like receptors. **Journal of Immunology**, v. 164, p. 966-72, 2000.

CASTLE, W.B. Grain-fed pigeons revisited: a pioneer test for vitamin B12. **British Journal of Experimental Pathology**, v. 66, p. 503-10, 1985.

CHEN, M. et al. Maternal temperature history activates Flowering Locus T in fruits to control progeny dormancy according to time of year. **Proceedings of the National Academy of Sciences**, v. 111, p. 18787-92.

CHEON, Y. et al. Immune-responsive gene 1 is a novel target of progesterone receptor and plays a critical role during Implantation in the mouse. **Endocrinology**, v. 144, n. 12, p. 5623-30, 2003.

CHESTER, E.M. et al. Optic atrophy in vitamin B12 deficiency in monkeys. **Acta Neurologica Scandinavica**, v. 61, p. 9-26, 1980.

COHN, M. The evolutionary context for a Self-Nonself discrimination. **Cellular and Molecular Life Sciences**, v. 67, n. 17, p. 2851-62, 2010.

CORDES, T., MICHELUTTI, A., HILLER, K. Itaconic acid: the surprising role of an industrial compound as a mammalian antimicrobial metabolite. **Annual Review of Nutrition**, v. 35, p. 12.1-12.23, 2015.

CORDES, T. et al. Immunoresponsive gene 1 and itaconate inhibit succinate dehydrogenase to modulate intracellular succinate levels. **Journal of Biological Chemistry**, v. 291, n. 27, p. 14274-84, 2016.

CROMBACH, A., HOGEWEG, P. Evolution of evolvability in gene regulatory networks. **PLoS Computational Biology**, v. 4, n. 7, e1000112, 2008.

DANCHIN, A., BRAHAM, S. Coenzyme B12 synthesis as a baseline to study metabolite contribution of animal microbiota. **Microbial Biotechnology**, v. 10, n. 4, p. 688-01, 2017

DARWIN, C.R. **On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life**. 5a ed. London: John Murray, 1869.

DAVIES, B., CHATTINGS, L.S., EDWARDS, S. Superoxide generation during phagocytosis by *Acanthamoeba castellanii*: similarities to the respiratory burst of immune phagocytes. **Journal of General Microbiology**, v. 137, p. 705-10, 1991.

DAVIS, M.M. et al. A murine T cell receptor gene complex: isolation, structure and rearrangement. **Immunological Reviews**, v. 81, p. 235-58, 1984.

DE DUVE, C. The onset of selection. **Nature**, v. 433, n. 7026, p. 581-2, 2005.

DEGRANDI, D. et al., The proinflammatory cytokine-induced IRG1 protein associates with mitochondria. **Journal of Interferon & Cytokine Research**, v. 29, n. 1, p. 55-67, 2009.

DEMIR-HILTON, E. Global distribution patterns of distinct clades of the photosynthetic picoeukaryote *Ostreococcus*. **The ISME Journal**, v. 5, n. 7, p. 1095-107, 2011.

DESJARDINS, M., HOUDE, M., GAGNON, E. Phagocytosis: the convoluted way from nutrition to adaptive immunity. **Immunological Reviews**, v. 207, p. 158-65, 2005.

DIACOVICH, L. et al. The infectious intracellular lifestyle of *Salmonella enterica* relies on the adaptation to nutritional conditions within the Salmonella-containing vacuole. **Virulence**, v. 8, n. 6, p. 975-92, 2017.

DIMROTH, P., EGGERER, H. Isolation of subunits of citrate lyase and characterization of their function in the enzyme complex. **Proceedings of the National Academy of Science**, v. 72, p. 3458-62, 1975.

DISCHER, D.E., MOONEY, D.J., ZANDSTRA, P.W. Growth factors, matrices, and forces combine and control stem cells. **Science**, v. 324, n. 5935, p. 1673-77, 2009.

DOOLITTLE, W.F. Lateral genomics. **Trends in Cell Biology**, v. 9, n. 12, p. M5-M8, 1999.

DOS REIS, M. et al. Uncertainty in the timing of origin of animals and the limits of precision in molecular timescales. **Current Biology**, v. 25, p. 2939-50, 2015.

DUNN, C.W. et al., Animal phylogeny and its evolutionary implications. **Annual Review of Ecology, Evolution, and Systematics**, v. 45, p. 371-95, 2014.

DUNN, M.F., RAMÍREZ-TRUJILLO, J.A., HERNÁNDEZ-LUCAS, I. Major roles of isocitrate lyase and malate synthase in bacterial and fungal pathogenesis. **Microbiology**, v. 155, n. 3, p. 3166-75, 2009.

DZIK, J.M. The ancestry and cumulative evolution of immune reactions. **Acta Biochimica Polonica**, v. 57, n. 4, p. 443-66, 2010.

EHRlich, P.R., RAVEN, P.H. Differentiation of populations: gene flow seems to be less important in speciation than the neo-darwinians thought. **Science**, v. 165, p. 1228-32, 1969.

EL-IMAM, A.A, DU, C. Fermentative itaconic acid production. **Journal of Biodiversity, Bioprospecting and Development**, v.1, n. 1, 2014.

EVANGELOPOULOU, G., et al. Animal salmonellosis: a brief review of “host adaptation and host specificity” of *Salmonella* spp. **Veterinary World**, v. 6, p. 703-08, 2013.

FANG, G. et al. Getting started in gene orthology and functional analysis. **PLoS Computational Biology**, v. 6, n. 3, 2010.

FELIX, M.A. Phenotypic evolution with and beyond genome evolution. **Current Topics on Developmental Biology**, v. 119, p. 291-347, 2016.

FELSENSTEIN, J. Confidence limits on phylogenies: an approach using the bootstrap. **Evolution**, v. 39, n. 4, p. 783-91, 1985.

FINN, R.D. et al. The Pfam protein families database: towards a more sustainable future. **Nucleic Acids Research**, v. 4, n. D1, 2016.

FITCH, W.M. Distinguishing homologous from analogous proteins. **Systematic Zoology**, v. 19, p. 99-106, 1970.

FLAJNIK, M.F., KASAHARA, M. Comparative genomics of the MHC: glimpses into the evolution of the adaptive immune system. **Immunity**, v. 15, p. 351-62, 2001.

FRANCO, I.S., SHUMAN, H.A., CHARPENTIER, X. The perplexing functions and surprising origins of *Legionella pneumophila* type IV secretion effectors. **Cell Microbiology**, v. 11, p. 1435-43, 2009.

FREEMAN, M. et al. An ancient, highly conserved family of cysteine-rich protein domains revealed by cloning type I and type II murine macrophage scavenger receptors. **Proceedings of the National Academy of Sciences**, v. 87, p. 8810-14, 1990.

FUMAGALLI, M., SIRONI, M., POZZOLI, U., FERRER-ADMETT-
LA, A., PATTINI, L., NIELSEN, R. Signature of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. **PLoS Genetics**, v. 7, n. 11, E1002355, 2011.

GAGNEUX, S. Host-pathogen coevolution in human tuberculosis. **Philosophical Transactions of the Royal Society B**, v. 367, p. 850-9, 2012.

GHANBARIAN, A.T., HURST, L.D. Neighboring genes show correlated evolution in gene expression. **Molecular Biology and Evolution**, v. 32, n. 7, p. 1748-66, 2015.

GREINER, J.W. et al. Recombinant interferon enhances monoclonal antibody-targeting of carcinoma lesions in vivo. **Science**, v. 235, p. 895-8, 1987.

GRISSA, I., VERGNAUD, G., POURCEL, C. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. **Nucleic Acids Research**, v. 35, p. W52-W57, 2007.

GUIRY, M.D., GUIRY, G.M. *AlgaeBase*. World-wide electronic publication. National University of Ireland Galway. Disponível em: <<http://www.algaebase.org>>. Acesso em: 20 setembro 2018.

HALL, C.J. et al. Immunoresponse gene 1 augments bactericidal activity of macrophage-lineage cells by regulating β -oxidation-dependent mitochondrial ROS production. **Cell Metabolism**, v. 18, n. 2, p. 265–78, 2013.

HAMILTON, W.D. Pathogens as causes of genetic diversity in their host populations. Em: ANDERSON, R.M., MAY, R.M. (Eds.) **Population Biology of Infectious Diseases**. 1a ed. Berlim: Springer, 1982. p. 269-96.

HARRIS, J.K. et al. The genetic core of the universal ancestor. **Genome Research**, v. 13, n. 3, p. 407-12, 2003.

HARTENSTEIN, V. The neuroendocrine system of invertebrates: a developmental evolutionary perspective. **Journal of Endocrinology**, v. 190, p. 555-70, 2006.

HASHIMOTO, C. The Toll gene of *Drosophila*, required for dorsal-ventral embryonic polarity, appears to encode a transmembrane protein. **Cell**, v. 52, p. 269-79, 1988.

HOLMES, E.C. What does virus evolution tell us about virus origins? **Journal of Virology**, v. 85, n. 11, p. 5247-51, 2011.

HORSWILL, A.R., ESCALANTE-SEMENERA, J.C. *Salmonella typhimurium* LT2 catabolizes propionate via the 2-methylcitric acid cycle. **Journal of Bacteriology**, v. 181, p. 5615-23, 1999.

HOZUMI, N., TONEGAWA, S. Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions. **Proceedings of the National Academy of Sciences**, v. 73, p. 3628-32, 1976.

HUERTA-CEPAS, J. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. **Nucleic Acids Research**, v. 44, D286–D293, 2016.

HUANG, S. et al. Genome analysis of the immune gene repertoire of amphioxus reveals extraordinary innate complexity and diversity. **Genome Research**, v. 18, p. 1112-36, 2008.

JAKLITSCH, W.M., KUBICEK, C.P., SCRUTTON, M.C. The subcellular organization of itaconate biosynthesis in *Aspergillus terreus*. **Journal of General Microbiology**, v. 137, n. 3, p. 533-9, 1991.

JANEWAY Jr., C.A. Approaching the asymptote? Evolution and revolution in immunology. **Cold Spring Harbor Symposia on Quantitative Biology**, v. 54, p. 1-13, 1989.

JANSEN, R. et al. Identification of genes that are associated with DNA repeats in prokaryotes. **Molecular Microbiology**, v. 43, n. 6, p. 1565-75, 2002.

KACSER, H., BURNS, J.A. The Molecular Basis of Dominance. **Genetics**, v. 97, n. 3-4, p. 639-66, 1981.

KANZOK, S.M. et al. Origin of Toll-like receptor-mediated innate immunity. **Journal of Molecular Evolution**, v. 58, p. 442-8, 2004.

KELLER, L.F., WALLER, D.M. Inbreeding effects on wild populations. **Trends in Ecology and Evolution**, v. 17, n. 5, p. 230-41, 2002.

KIMURA, M., Evolutionary rate at the molecular level. **Nature**, v. 217, p. 624-6, 1968.

KING, HITTINGER, CARROLL, Evolution of key cell signaling and adhesion protein families predates animal origins. **Science**, v. 18, n. 301, p. 361-3, 2003.

KINOSHITA, 1931 *apud* BENTLEY, R., THIESSEN, C.P. Biosynthesis of itaconic acid in *Aspergillus terreus*. I. Tracer studies with C14-labeled substrates. **Journal of Biological Chemistry**, v. 226, n. 2, p. 673-87, 1957.

KLEMENT, T., BÜCHS, J. Itaconic acid--a biotechnological process in change. **Bioresource Technology**, v.135, p. 422-31, 2013.

KOONIN, E.V. How many genes can make a cell: the minimal gene-set concept. **Annu. Rev. Genomics Hum. Genet**, v. 1, p. 99-116, 2000.

KOONIN, E.V., MARTIN, W. On the origin of genomes and cells within inorganic compartments. **Trends in Genetics**, v. 21, n. 12, p. 647-654, 2005.

KOONIN, E.V. Orthologs, paralogs and evolutionary genomics. **Annual Review of Genetics**, v. 39, p. 309-38, 2015.

KRAUSE, C.D., PESTKA, S. Evolution of the class 2 cytokines and receptors, and discovery of new friends and relatives. **Pharmacology and Therapeutics**, v. 106, p. 299-346, 2005.

KUMAR, A., PLACONE, J.K., ENGLER, A.J. Understanding the extracellular forces that determine cell fate and maintenance.

Development, v. 144, p. 4261-70, 2017.

LALAND, K.N. et al. The extended evolutionary synthesis: its structure, assumptions and predictions. **Proceedings of the Royal Society B**, v.

282, n. 1813, p. 1-14, 2015.

LAMPROPOULOU, V. et al. Itaconate links inhibition of succinate dehydrogenase with macrophage metabolic remodeling and regulation of inflammation. **Cell Metabolism**, v. 24, p. 158-66, 2016.

LANDE, R. Genetics and demography in biological conservation.

Science, v. 241, p. 1455-60, 1988.

LEE, C.G.L. et al. Cloning and analysis of gene regulation of a novel

LPS-inducible cDNA. **Immunogenetics**, v. 41, p. 263-70, 1995.

LEIPPE, M. et al. Amoebapores, a family of membranolytic peptides from cytoplasmic granules of *Entamoeba histolytica*: isolation, primary structure, and pore formation in bacterial cytoplasmic membranes.

Molecular Microbiology, v. 14, p. 895-904, 1994.

LEMAITRE, B. et al. The dorsoventral regulatory gene cassette spätzle/Toll/cactus controls the potent antifungal response in

Drosophila. **Cell**, v. 86, p. 973-83, 1996.

LI, H. et al. Different neurotropic pathogens elicit neurotoxic CCR9- or neurosupportive CXCR3-expressing microglia. **Journal of**

Immunology, n. 177, p. 3644-56, 2006.

LIONGUE, C., SERTORI, R., WARD, A.C. Evolution of cytokine receptor signaling. **The Journal of Immunology**, v. 197, n. 1, p. 11-8,

2016.

LOMBARD, J., LÓPEZ-GARCIA, P., MOREIRA, D. Phylogenomic investigation of phospholipid synthesis in archaea. **Archaea**, v. 2012, n. 630910, 2012.

LYNCH, M., CONNERY, J.S. The evolutionary fate and consequences of duplicate genes. **Science**, v. 290, p. 1151-5, 2000.

LYNCH, M., FORCE, A. The probability of duplicate gene preservation by subfunctionalization. **Genetics**, v. 154, n. 1, p. 459-73, 2000.

MACQUEEN, J.B. Some methods for classification and analysis of multivariate observations. **Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability**. Berkeley: University of California Press, v. 1, p. 281-97, 1967.

MANSUR, D.S., SMITH, G.L., FERGUSON, B.J. Intracellular sensing of viral DNA by the innate immune system. **Microbes and Infection**, v. 16, n. 12, p. 1002-12, 2014.

MARCHALONIS, J.J. et al. Phylogenetic emergence and molecular evolution of the immunoglobulin family. **Advances in Immunology**, v. 70, p. 417-506, 1998.

MARTIN, W. et al. Hydrothermal vents and the origin of life. **Nature Reviews Microbiology**, v. 6, p. 805-14, 2008.

MCFADDEN, B.A., PUROHIT, S. Itaconate, an isocitrate lyase-directed inhibitor in *Pseudomonas indigofera*. **Journal of Bacteriology**, v. 131, n. 1, p. 136-44, 1977.

MEDZHITOV, R., JANEWAY Jr., C. Innate immune recognition: mechanisms and pathways, **Immunological Reviews**, v. 173, p. 89-97, 2000.

METCHNIKOFF, E. **Immunity in Infective Diseases**. Cambridge: University Press, 1905.

MEYER, A., SCHARTL, M. Gene and genome duplication in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. **Current Opinion in Cell Biology**, v. 11, p. 699-704, 1999.

MICHELUTTI, A. et al. Immune-responsive gene 1 protein links metabolism to immunity by catalyzing itaconic acid production. **Proceeding of the National Academy of Sciences**, v. 110, n. 19, p. 7620-25, 2013.

MORA, C. et al. How many species are there on Earth and in the ocean? **PloS Biology**, v. 9, n. 8, 2011.

MUKHERJEE, S. et al. Genomes OnLine Database (GOLD) v.6: data updates and feature enhancements. **Nucleic Acids Research**, v. 43, n. 45, p. D446-D456, 2016.

MUKHERJEE, K., KORITHOSKI, B., KOLACZKOWSKI, B. Ancient origins of vertebrate-specific innate antiviral immunity. **Molecular Biology and Evolution**, v. 31, n. 1, p. 140-53, 2014.

MUÑOZ-ELÍAS, E.J. *Mycobacterium tuberculosis* isocitrate lyase 1 and 2 are jointly required for in vivo growth and virulence. **Nature Medicine**, v. 11, n. 6, p. 638-44, 2006.

MURAMATSU, M. et al. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. **Cell**, v. 102, p. 553-63, 2000.

MUTA, T., IWANAGA, S. The role of hemolymph coagulation in innate immunity. **Current Opinion in Immunology**, v. 8, p. 41-7, 1996.

NAIR, S. Irg1 expression in myeloid cells prevents immunopathology during *M. tuberculosis* infection. **Journal of Experimental Medicine**, v. 215, n. 4, p. 1035-45, 2018.

NAPPI, A.J., OTTAVIANI, E. Cytotoxicity and cytotoxic molecules in invertebrates. **Bioessays**, v. 22, p. 469-90, 2000.

NAUJOKS, J. et al. IFNs modify the proteome of *Legionella*-containing vacuoles and restrict infection via IRG1-derived itaconic acid. **PLoS Pathogens**, v. 12, n. 2, 2016.

OETTINGER, M.A. et al. RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. **Science**, v. 248, p. 1517-23, 1990.

OHNO, S. **Evolution by Gene Duplication**. 1a ed. New York: Springer, 1970.

ORGAN, C.L., MORENO, R.G., EDWARDS, S.V. Three tiers of genome evolution in reptiles. **Integrative and Comparative Biology**, v. 48, n. 4, p. 494-504, 2008.

ORGEL, L.E. Prebiotic chemistry and the origin of the RNA World. **Critical Reviews of Biochemistry and Molecular Biology**, v. 39, p. 99-123, 2004.

ORTALDO, J.R. et al. Effects of recombinant and hybrid recombinant human leukocyte interferons on cytotoxic activity of natural killer cells. **Journal of Biological Chemistry**, v. 258, p. 15011-5, 1983.

PADOVANI, D., BANERJEE, R. Alternative pathways for radical dissipation in an active site mutant of B12-dependent methylmalonyl-CoA mutase. **Biochemistry**, v. 45, n. 9, p. 2951-9, 2006.

PANCER, Z. Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey. **Nature**, v. 430, p. 174-80, 2004.

PARFREY, L.W. et al. Estimating the time of early eukaryotic diversification with multigene molecular clocks. **Proceeding of the National Academy of Sciences**, v. 108, n. 33, p. 13624-13629, 2011.

PARIA, B.C., HUET-HUDSON, Y.M., DEY, S.K. Blastocyst's state of activity de- termines the “window” of implantation in the receptive mouse uterus. **Proceedings of the National Academy of Science**, v. 90, p. 10159-62, 1993.

PARTE, A.C. LPSN—List of Prokaryotic names with Standing Nomenclature. **Nucleic Acids Research**, v. 42, n. D1, p. D613–D616, 2014.

PEARSON, W.R. An introduction to sequence similarity (“homology”) searching. **Current Protocols in Bioinformatics**, n. 3, 2013.

PREUSSE, M. et al. Infection- and procedure-dependent effects on pulmonary gene expression in the early phase of influenza A virus infection in mice. **BioMed Central Microbiology**, v. 13, 2011.

PSYCHOYOS, A. Uterine receptivity for nidation. **Annals of the New York Academy of Science**, v. 476, p. 36-42.

PSYCHOYOS, A. Hormonal control of uterine receptivity for nidation. **Journal of Reproduction and Fertility**, v. 25, p. 17-28, 1976.

RAST, J.P. et al. α , β , γ and δ T cell antigen receptor genes arose early in vertebrate phylogeny. **Immunity**, v. 6, p. 1-11, 1997.

RAST, J.P. et al. Genomic insights into the immune system of the sea urchin. **Science**, v. 314, p. 952-6, 2006.

ROSENBERG, E., ZILBER-ROSENBERG, I. Microbes drive evolution of animals and plants: the hologenome concept. **MBio**, v. 7, n. 2, e01395, 2016.

RUGGIERO, M.A. A higher level classification of all living animals. **PLoS ONE**, v. 10, n. 4, p. 1-60, 2015.

RUSSEL, M.J., DANIEL, R.M., HALL, A.J. On the emergence of life via catalytic iron-sulphide membranes. **Terra Nova**, v. 5, p. 343-7, 1993.

RUSSEL, M.J. et al. A hydrothermally precipitated catalytic iron sulphide membrane as a first step toward life. **Journal of Molecular Evolution**, v. 39, p. 231-43, 1994.

RUSSEL M.J., HALL, A.J. The emergence of life from iron monosulphide bubbles at a submarine hydrothermal redox and pH front. **Journal of the Geological Society of London**, v. 154, p. 377-402, 1997.

SACCHERI, I., HANSKI, I. Natural selection and population dynamics. **Trends in Ecology and Evolution**, v. 21, n. 6, p. 341-47, 2006.

SAKAI, A. et al. Itaconate reduces visceral fat by inhibiting fructose-2,6-biphosphate synthesis in rat liver. **Nutrition**, v. 20, n. 11, p. 997-1002, 2004.

SASIKARAN, J., et al. Bacterial itaconate degradation promotes pathogenicity. **Nature Chemical Biology**, v. 10, p. 371-77, 2014.

SATOH, N., ROKHSAR, D., NISHIKAWA, T. Chordate evolution and the three-phylum system. **Proceedings of the Royal Society B**, v. 281, p. 1-10, 2014.

SHEN, H. et al. The human knockout gene CLYBL connects itaconate to vitamin B12. **Cell**, v. 171, p. 1-12, 2017.

SCHLUTER, S.F. et al. 'Big Bang' emergence of the combinatorial immune system. **Developmental and Comparative Immunology**, v. 23, p. 107-11, 1999.

SEGOVIA, R., TAM, A.S., STIRLING, P.C. Dissecting genetic and environmental mutation signatures with model organisms. **Trends in Genetics**, v. 31, p. 465-74, 2015.

SIEVERS et al., Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. **Molecular Systems Biology**, v. 7, n. 539, 2011.

SMITH, J. et al. Systems analysis of immune responses in Marek's disease virus-infected chickens identifies a gene involved in susceptibility and highlights a possible novel pathogenicity mechanism. **Journal of Virology**, v. 85, n. 21, p. 11146-58, 2011.

SMITH, N.H. et al. Myths and misconceptions: the origin and evolution of *Mycobacterium tuberculosis*. **Nature Reviews Microbiology**, v. 7, p. 537-44, 2009.

SOAVE, O., BRAND, C.D. Coprophagy in animals: a review. **Cornell Veterinarian**, v. 81, p. 357-64, 1991.

SOBER, E. **From a biological point of view: essays in evolutionary philosophy**. Cambridge: Cambridge University Press, p. 28-49, 1994.

SÖDERHÄLL, K., CERENIUS, L., JOHANSSON, M.W. The prophenoloxidase activating system and its role in invertebrate defense. **Annals of the New York Academy of Sciences**, v. 712, p. 155-61, 1994.

SÖDING, J. Protein homology detection by HMM–HMM comparison.v. **Bioinformatics**, v. 21, n. 7, p. 951-60, 2005.

SOKAL, R.R., MICHENER, C.D. A statistical method for evaluating systematic relationships. **University of Kansas Science Bulletin**, v. 28, p. 1409-38, 1958.

STEIGER, M.G. et al. Biochemistry of microbial itaconic acid production. **Frontiers in Microbiology**, v. 4, n. 23, 2013.

STRELKO, C.L., et al. Itaconic acid is a mammalian metabolite induced during macrophage activation. **Journal of the American Chemical Society**, v. 133, n. 41, p. 16386-9, 2011.

SUGIMOTO, M., et al. Non-targeted metabolite profiling in activated macrophage secretion. **Metabolomics**, v. 8, n. 4, p. 24-633, 2012.

SUZUKI, R., SHIMODAIRA, H. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. **Bioinformatics**, v. 22, n. 12, p. 1540-2, 2006.

TANNAHILL, G.M. et al. Succinate is an inflammatory signal that induces IL-1 β through HIF-1 α . **Nature**, v. 496, n. 7444, p. 238-42, 2013.

VORZIMMER, P.J. Charles Darwin and blending inheritance. **IBIS**, v. 54, n. 177, p. 371-90, 1963.

WADDELL, D.R., DUFFY, K.T. Breakdown of self/non-self recognition in cannibalistic strains of the predatory slime mold, *Dictyostelium caveatum*. **Journal of Cell Biology**, v. 102, p. 298-305, 1986.

WAGNER, G.P., PAVLICEV, M., CHEVERUD, J.M. The road to modularity. **Nature Reviews Genetics**, v. 8, n. 12, p. 921-31, 2007.

WANG, S.F., ADLER, J., LARDY, H.A. The pathway of itaconate metabolism by liver mitochondria. **The Journal of Biological Chemistry**, v. 236, p. 26-30, 1961.

WARREN, W.C. et al. Genome analysis of the platypus reveals unique signatures of evolution. **Nature**, v. 453, n. 7192, p. 175-83, 2008.

WEST-EBERHARD, M.J. **Developmental plasticity and evolution**. 1a ed. New York: Oxford University Press, 2003. p. 16.

WILLIAMS, T.A. et al. An archaeal origin of eukaryotes supports only two primary domains of life. **Nature**, v. 504, p. 231-36, 2013.

WILKINSON, H.A., FITZGERALD, K., GREENWALD, I. Reciprocal changes in expression of the receptor lin-12 and its ligand lag-2 prior to commitment in a *C. elegans* cell fate decision. **Cell**, v. 79, n. 7, p. 1187-98, 1994.

WOESE, C.R., FOX, G.E. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. **Proceedings of the National Academy of Sciences**, v. 74, n. 11, p. 5088-90, 1977.

YAHIRO, K. et al. Breeding of *Aspergillus terreus* mutant TN-484 for an itaconic acid production with high yield. **Journal of Fermentation and Bioengineering**, v. 79, n. 5, p. 506-8, 1995.

YANG, Z., RANNALA, B. Molecular phylogenetics: principles and practices. **Nature Reviews Genetics**, v. 13, n. 5, p. 303-14, 2012.

YUEN, B., BAYES, J.M., DEGNAN, S.M. The characterization of sponge NLRs provides insight into the immune gene family in animals. **Molecular Biology and Evolution**, v. 31, n. 1, p. 106-20, 2013.

ZUCKERKANDL, E., PAULING, L. Evolutionary divergence and convergence in proteins. Em: BRYSON, V., VOGEL, H.J. (Eds.) **Evolving Genes and Proteins**. 1a ed. New York: Academic Press, 1965. p. 97-166.

APÊNDICES

Devido às dimensões de alguns arquivos incluídos como apêndices, os mesmos podem ser encontrados acompanhando este documento sob o diretório “Apêndices”. Adiante seguem as legendas dos respectivos documentos.

APÊNDICE A – Sequências eggNOG

Arquivo externo

Quadro contendo os dados sobre as sequências consideradas na construção da Figura 3, a árvore filogenética da enzima cis-aconitase descarboxilase (CAD) nos três domínios da vida. Os códigos de acesso e os nomes científicos de cada espécie foram extraídos diretamente da página da família de proteínas MmgE/PrpD, ancorada na base de dados eggNOG e disponível no seguinte endereço: <<https://goo.gl/taFnnW>>. A classificação taxonômica de cada espécie foi coletada manualmente a partir de uma série de bases de dados, listadas a seguir. A escolha da nomenclatura taxonômica frente aos principais casos de divergência de opiniões na literatura está justificada na seção 2.4, em Metodologia.

Algaebase (<<http://www.algaebase.org/>>)

Encyclopedia of Life (<<http://www.eol.org>>)

FishBase (<<https://goo.gl/9ExYUx>>)

The International Union for Conservation of Nature (IUCN) list of threatened species [do inglês, Lista de espécies ameaçadas da IUCN (<<http://iucnredlist.org/>>)]

Mycobank (<<http://www.mycobank.org/>>)

National Center for Biotechnological Information (NCBI) (<<https://goo.gl/6qdEZ4>>)

Tree of Life (<<http://tolweb.org/tree/>>)

UniProt (<<http://uniprot.org/>>)

APÊNDICE B – Linhas de comando

Linhas de comando chave empregadas na geração das árvores filogenéticas (Figuras 3 e 6), do diagrama de Venn (Figura 7A) e do *heatmap* (Figura 7B).

Árvores filogenéticas

```
## Alinhamento das sequências com Clustal Omega (shell)
clustalo --infile ARQUIVO.fasta --threads 4 --MAC-RAM 4000
--verbose --outfmt clustal --outfile ARQUIVO.clustalo --output-
order tree-order --seqtype protein --distmat-out ARQUIVO.pim
--percent-id --guidetree-out ARQUIVO.tree --full --force
```

Bootstrapping ®

Carregamento dos pacotes

```
library(ape)
library(pvclust)
```

Criação de um objeto a partir da matriz de distâncias resultante do alinhamento

```
OBJETO<-read.table ("ARQUIVO.pim2", header=FALSE, sep=" ")
```

Definição do OBJETO como matriz

```
OBJETO<-as.matrix(cad)
```

Atribuição dos elementos contidos na primeira coluna do OBJETO como nomes das linhas

```
rownames(OBJETO)<-OBJETO[,1]
```

Remoção da primeira coluna

```
OBJETO<-OBJETO[,-1]
```

Atribuição dos nomes das linhas aos nomes das colunas

```
colnames(OBJETO)<-rownames(OBJETO)
```

Alteração do tipo de matriz, de dados para numérica

```
OBJETO<-(apply(OBJETO[,], 2, as.numeric))
```

Análise filogenética com o pacote Pvclust

```
AF.OBJETO<-pvclust(OBJETO)
```

Diagrama de Venn (R)

Carregamento do pacote

```
library(venn)
```

```
# Geração do diagrama
venn(tab01.venn, lty=2, cexil = 0.95, cexsn = 1.1,
col=c("thistle","tan1"), zcolor=c("thistle","tan1"),
opacity=0.4)
```

```
### Heatmap
```

```
## Carregamento do pacote
library(gplots)
```

```
# Criação da paleta de cores
cor=colorRampPalette(c("white", "palegreen3", "palegreen4"))
```

```
# Geração do heatmap
heatmap.2(tabela, Rowv=NULL, Colv=NULL, dendrogram=NULL,
col=cor, scale="none", trace="none", key=FALSE, colsep=c(1:29),
rowsep=c(1:29), sepwidth=c(0.01,0.01))
```

APÊNDICE C – Árvore filogenética da CAD nos três domínios

Arquivo externo

Árvore filogenética original da CAD nos três domínios (relativo à figura 3). As sequências foram alinhadas com o alinhador Clustal Omega e os valores de confiabilidade foram gerados com o pacote Pvclust.

Apêndice D – Tabela de valores estatísticos 1
Árvore filogenética da CAD nos três domínios
Arquivo externo

Tabela contendo os valores de bootstrap (BP) e *approximately unbiased* (AU) calculados pelo pacote Pvcust, executado no ambiente R para a árvore filogenética da CAD nos três domínios da vida (Figura 3).

APÊNDICE E – Análise de sítios ativos

Análise dos parâmetros dos alinhamentos e da presença dos sítios ativos preditos para a CAD referentes às sequências resultantes da busca com a BLAST. Todas as sequências dos respectivos filós foram descarregadas do GenBank a partir de seus códigos de acesso e alinhadas com o Clustal Omega à proteína humana e/ou intrafilo. O alinhamento foi exportado para a página da ferramenta *online* ESPrift (<<http://esprift.ibcp.fr/ESPrift/ESPrift/>>), a partir da qual foram extraídos os fragmentos de imagem que seguem.

Stylophora pistillata (XP_022788243.1)	54	RLKLAL	ELV...	ALGFLT	M...	TVLF	ITWQIT	KKRP	DA	A	V	QNP																													
Stylophora pistillata (PFX27052.1)	34	RLKLAL	ELV...	ALGFLT	M...	TVLF	ITWQIT	KKRP	DA	A	V	QNP																													
Drosophila ficusphila (XP_017045452.1)	28	VAAD	PVPV	EA	EA	EM	VI	NR	IDNA																												
Drosophila obscura (XP_022212151.1)	28	VAAD	PVPV	EA	EA	EM	VI	NR	IDNA																												
Drosophila obscura (XP_022234573.1)	28	VAAD	PVPV	EA	EA	EM	VI	NR	IDNA																												
Drosophila serrata	1																												
Drosophila eugracilis	1																												
Drosophila elegans	28	VAAD	PVPV	EA	EA	EM	VI	NR	IDNA																												
Drosophila ficusphila (XP_017045444.1)	28	VAAD	PVPV	EA	EA	EM	VI	NR	IDNA																												
Drosophila kikkawai	28	VAAD	PVPV	EA	EA	EM	VI	NR	IDNA																												
Drosophila bipectinata	28	VAAD	PVPV	EA	EA	EM	VI	NR	IDNA																												
Schistosoma haematobium	106	RRKRG	YRPL	TD	LN	SR	IN	IP	KAR	V	LT	D	KE	LSE	...	LR	SK	T	...																						
Opisthorchis viverrini	292	...	PK	DE	GL	GL	T	V	GV	Y	R	D	P	A	K	S	K	L	H	L	G	G	E	Y	...	T	S	G	I	F	V	...									
Stylophora pistillata (PFX34234.1)	3415	RRK	V	D	K	K	Y	Q	E	N	L	F	A	Y	V	D	K	T	R	A	E	K	F	V	N	L	E	D	L	S	E	E	Q	F	S	S	P	I	D	...	
Stylophora pistillata (XP_022787984.1)			
Nematostella vectensis			
Homo sapiens	68	VWGQ	P	D	I	R	L	P	P	T	Y	A	A	F	V	N	G	V	A	I	H	S	M	D	F	D	D	T	W	H	P	A	T	H	P	S	G	A	V	L	...
Branchiostoma	62	LLGR	K	D	R	T	T	S	A	P	L	A	A	F	A	N	G	V	S	A	H	S	M	D	F	D	D	T	W	H	P	A	T	H	P	S	A	P	V	L	...
Branchiostoma belcheri	87	LLGR	K	D	R	A	T	S	A	P	L	A	A	F	A	N	G	V	S	A	H	S	M	D	F	D	D	T	W	H	P	A	T	H	P	S	A	P	V	L	...
Mizuhopecten yessoensis (XP_021359564.1)	127	IWGT	G	G	L	K	A	S	M	A	M	A	A	Y	L	N	G	A	S	C	H	A	M	D	F	D	D	T	W	H	P	A	T	H	P	S	G	P	V	L	...
Mizuhopecten yessoensis (XP_021359565.1)	118	IWGT	G	G	L	K	A	S	M	A	M	A	A	Y	L	N	G	A	S	C	H	A	M	D	F	D	D	T	W	H	P	A	T	H	P	S	G	P	V	L	...
Mizuhopecten yessoensis (XP_021359566.1)	115	IWGT	G	G	L	K	A	S	M	A	M	A	A	Y	L	N	G	A	S	C	H	A	M	D	F	D	D	T	W	H	P	A	T	H	P	S	G	P	V	L	...
Ostrea edulis			
Crassostrea virginica (XP_022312773.1)	102	IWGT	Q	N	R	R	T	G	P	G	M	A	A	F	I	N	G	S	N	C	H	S	M	D	F	D	D	T	W	H	P	A	T	H	P	S	S	P	V	L	...
Crassostrea virginica (XP_022319601.1)	102	IWGT	Q	N	R	R	T	G	P	G	M	A	A	F	I	N	G	S	N	C	H	S	M	D	F	D	D	T	W	H	P	A	T	H	P	S	S	P	V	L	...
Crassostrea gigas (XP_011438844.1)	96	IWGT	S	N	K	R	T	R	P	G	M	A	A	F	I	N	G	S	N	C	H	S	M	D	F	D	D	T	W	H	P	A	T	H	P	S	S	P	I	L	...
Crassostrea gigas (EKC41910.1)	117	IWGT	S	N	K	R	T	R	P	G	M	A	A	F	I	N	G	S	N	C	H	S	M	D	F	D	D	T	W	H	P	A	T	H	P	S	S	P	I	L	...
Octopus bimaculoides	129	IWGS	S	N	K	R	A	P	I	G	L	A	T	Y	L	N	G	I	S	I	H	S	M	D	F	D	D	T	W	H	P	A	T	H	P	S	G	P	V	L	...
Stylocryptus labrata	158	IWGS	N	R	M	R	A	P	A	T	A	Y	I	N	G	S	I	H	S	M	D	F	D	D	T	W	H	P	A	T	H	P	S	G	P	V	L	...			
Lottia gigantea	132	LWGS	N	K	R	K	V	A	P	A	I	A	Y	I	N	G	V	N	V	H	S	M	D	F	D	D	T	W	H	P	A	T	H	P	S	G	P	V	L	...	
Pomacea canaliculata	119	IWGS	N	G	K	T	A	S	P	P	I	A	A	F	I	N	G	S	I	V	H	S	M	D	F	D	D	T	W	D	F	A	T	H	P	S	G	P	V	L	...



Stylophora pistillata (XP_022788243.1)	263	N.....TYVLEQLS.....	K
Stylophora pistillata (PFX27052.1)	196	N.....TYVLEQLS.....	K
Drosophila ficusphila (XP_017045452.1)	157	..RAICLHKHKIDHVAH.....	L
Drosophila obscura (XP_022212151.1)	157	..RAICLHKHKIDHVAH.....	L
Drosophila obscura (XP_022234573.1)	157	..RAICLHKHKIDHVAH.....	L
Drosophila serrata	116	..RAICLHKHKIDHVAH.....	L
Drosophila eugracilis	116	..RAICLHKHKIDHVAH.....	L
Drosophila elegans	157	..RAICLHKHKIDHVAH.....	L
Drosophila ficusphila (XP_017045444.1)	157	..RAICLHKHKIDHVAH.....	L
Drosophila kikkawai	157	..RAICLHKHKIDHVAH.....	L
Drosophila bipectinata	157	..RAICLHKHKIDHVAH.....	L
Schistosoma haematobium	251	VAPGSKVAGLRMLDQPYMDLIEAAAMAYARDLIDIYSASWGPT	
Opisthorchis viverrini	397	ASHVR.CTEFRTW.....H.....GQKQESTRIPLYHTSLGAA	
Stylophora pistillata (PFX34234.1)	3530	.APSR.VINFNN.....	I
Stylophora pistillata (XP_022787984.1)		I
Nematostella vectensis		I
Homo sapiens	145	..LHF.AKEAND.....	M
Branchiostoma	139	..LRT.SAQAKN.....	N
Branchiostoma belcheri	164	..LRT.SAQAKN.....	I
Mizuhopecten yessoensis (XP_021359564.1)	204	..LNC.STQARN.....	I
Mizuhopecten yessoensis (XP_021359565.1)	195	..LNC.STQARN.....	I
Mizuhopecten yessoensis (XP_021359566.1)	192	..LNC.STQARN.....	I
Ostrea edulis		I
Crassostrea virginica (XP_022312773.1)	179	..LKC.SLQSKN.....	I
Crassostrea virginica (XP_022319601.1)	179	..LKC.SLQSKN.....	I
Crassostrea gigas (XP_011438844.1)	173	..LRC.SLQSNM.....	I
Crassostrea gigas (EKC41910.1)	194	..LRC.SLQSNM.....	I
Octopus bimaculoides	206	..LMS.SAKSKN.....	I
Blomphalaria glabrata	235	..LRS.SASAKS.....	I
Lottia gigantea	209	..LRC.SEIAKE.....	I
Pomacea canaliculata	196	..LRC.SDSARA.....	I



Stylophora pistillata (XP_022788243.1)	273	TELKSPFCRIAPHY.....	I
Stylophora pistillata (PFX27052.1)	206	TELKSPFCRIAPHY.RNPNGSTKSNELIFFSNLGYSLIARIFAE	
Drosophila ficusphila (XP_017045452.1)	184	..LKLPPPAVITYQAIG.....QALH.....	V
Drosophila obscura (XP_022212151.1)	184	..LKLPPPAVITYQAIG.....QALH.....	V
Drosophila obscura (XP_022234573.1)	184	..LKLPPPAVITYQAIG.....QALH.....	V
Drosophila serrata	143	..LRLPPPAVITYQAIG.....QALH.....	V
Drosophila eugracilis	143	..LRLPPPAVITYQAIG.....QALH.....	V
Drosophila elegans	184	..MRLPPPAVITYQAIG.....HADH.....	V
Drosophila ficusphila (XP_017045444.1)	184	..LRLPPPAVITYQAIG.....QALH.....	V
Drosophila kikkawai	184	..MRLPPPAVITYQAIG.....QALH.....	V
Drosophila bipectinata	184	..MRLPPPAVITYQAIG.....QALH.....	V
Schistosoma haematobium	295DDGTITVDGPRNQT.....	M
Opisthorchis viverrini	429	TADKPDPEHVGPGFTGLAGPTSRDLIQ.....QSGLN	
Stylophora pistillata (PFX34234.1)	3541	TQLSSPPQEDRTSGSLDGQS.....LAQ.....	SEWLS
Stylophora pistillata (XP_022787984.1)		
Nematostella vectensis		
Homo sapiens	155	PKRFHPPSPVGTIGSAAAASKFLG..LSS.....	TKCR
Branchiostoma	149	PWRFHPPAVVGVMGSAACAKLQ..HDP.....	TRVR
Branchiostoma belcheri	174	PWRFHPPAVVGVMGSAACAKLQ..HDP.....	VHVR
Mizuhopecten yessoensis (XP_021359564.1)	214	PHRLHPPAIVGVMGSAACSRLLG..HDP.....	QKCR
Mizuhopecten yessoensis (XP_021359565.1)	205	PHRLHPPAIVGVMGSAACSRLLG..HDP.....	QKCR
Mizuhopecten yessoensis (XP_021359566.1)	202	PHRLHPPAIVGVMGSAACSRLLG..HDP.....	QKCR
Ostrea edulis		
Crassostrea virginica (XP_022312773.1)	189	PNRLHPPAIVGVMGSAASANLQ..LNS.....	QKSL
Crassostrea virginica (XP_022319601.1)	189	PNRLHPPAIVGVMGSAASANLQ..LNS.....	QKSL
Crassostrea gigas (XP_011438844.1)	183	PNRLHPPAIVGVMGSAASANLQ..LEP.....	KKCL
Crassostrea gigas (EKC41910.1)	204	PNRLHPPAIVGVMGSAASANLQ..LEP.....	KKCL
Octopus bimaculoides	216	PERFHPPAVVGVMGAAATARLQ..LGP.....	LKTK
Blomphalaria glabrata	245	PCRFHPPAVVGVMGSAACANLQ..LGP.....	EKCR
Lottia gigantea	219	PKRFHPPAVVGVMGSAASANLQ..LGP.....	EKCR
Pomacea canaliculata	206	PDRFHPPAVVGVMGSAACARLQ..AGP.....	STVR



Stylophora pistillata (XP_022788243.1)	300	KFANSDYEGWVER.....	[L][I][I][I][I]	NIFQPLGMG
Stylophora pistillata (PFX27052.1)	249	KFANSDYEGWVER.....	[L][I][I][I][I]	NIFQPLGMG
Drosophila fucusphila (XP_017045452.1)	202	TTATRQSRKGEIS..	..	SWKAFAPAHAG
Drosophila obscura (XP_022212151.1)	202	TTATRQSRKGEIS..	..	SWKAFAPAHAG
Drosophila obscura (XP_022234573.1)	202	TTATRQSRKGEIS..	..	SWKAFAPAHAG
Drosophila serrata	161	TTATRQSRKGEIS..	..	SWKAYAPAHAG
Drosophila eugracilis	161	TTATRQSRKGEIS..	..	SWKAYAPAHAG
Drosophila elegans	202	TTATRQSRKGEIS..	..	SWKAYAPAHAS
Drosophila fucusphila (XP_017045444.1)	202	TTATRQSRKGGIS..	..	SWKAYAPAHAS
Drosophila kikkawai	202	TTATRQSRKGEIS..	..	SWKAYAPAHAS
Drosophila bipectinata	202	TTATRQSRKGEIS..	..	SWKAYAPAHAS
Schistosoma haematobium	309	RALVDGVNKGKRGKGLYVWASGDGG		
Opisthorchis viverrini	463	SHS...RPSIGQMPTIANGASR..TRQNLQKVITAYVTYAKRS..			
Stylophora pistillata (PFX34234.1)	3573	TEKSDAGCYIGRVSARPSAQKRRQKRPQGNRRKFFVFFVSPDG..			
Stylophora pistillata (XP_022787984.1)				
Nematostella vectensis				
Homo sapiens	186	.EALAIIVSHAGAPMANAATQT	KPLHIGNAAKHGI..		
Branchiostoma	180	.NAMSAAASYAGAPMANAGTT	AKPLHAGNAARFGL..		
Branchiostoma belcheri	205	.NALSAASASYAGAPMANAGTT	AKPLHAGNAARFGL..		
Mizuhopecten yessoensis (XP_021359564.1)	245	.HALAIAASFAGAPMANAGTT	TKPLHAGKSARFGL..		
Mizuhopecten yessoensis (XP_021359565.1)	236	.HALAIAASFAGAPMANAGTT	TKPLHAGKSARFGL..		
Mizuhopecten yessoensis (XP_021359566.1)	233	.HALAIAASFAGAPMANAGTT	TKPLHAGKSARFGL..		
Ostrea edulis				
Crassostrea virginica (XP_022312773.1)	220	.HALAIAASFAGAPMANAGTT	LTKPLHAGKSARSGL..		
Crassostrea virginica (XP_022319601.1)	220	.HALAIAASFAGAPMANAGTT	LTKPLHAGKSARSGL..		
Crassostrea gigas (XP_011438844.1)	214	.HALSIAASFAGAPMANAGTT	LAKPLHAGKSARCGL..		
Crassostrea gigas (EKC41910.1)	235	.HALSIAASFAGAPMANAGTT	LAKPLHAGKSARCGL..		
Octopus bimaculoides	247	.HALAIAASFAGAPMANAGTT	CKPLHAGKSARFGL..		
Biomphalaria glabrata	276	.NALGIAASFAGAPMANAGTT	KKPLHAGKAARFGL..		
Lottia gigantea	250	.HALAIAASFAGAPMANAGTT	KKPLHSGKAARFGL..		
Pomacea canaliculata	237	.HALAVATSFAGAPMANAGTT	PKPVHAGKAARFGM..		



Stylophora pistillata (XP_022788243.1)	419	..TKSGFVPGYSA.....	GFAP	[P][E][F][K]	GAIALVS
Stylophora pistillata (PFX27052.1)	368	..TKSGFVPGYSA.....	GFAP	[P][E][F][K]	GAIALVS
Drosophila fucusphila (XP_017045452.1)				
Drosophila obscura (XP_022212151.1)	333	.DPQKFDPT...A..SRETLDH..	SIMY	[I][V][A][V][A][L]
Drosophila obscura (XP_022234573.1)	333	.DPQKFDPT...A..SRETLDH..	SIMY	[I][V][A][V][A][L]
Drosophila serrata	292	.DPQKFDPH...A..SRETLDH..	SIMY	[I][V][A][V][A][L]
Drosophila eugracilis	292	.DPQKFDPH...A..SRETLDH..	SIMY	[I][V][A][V][A][L]
Drosophila elegans	333	.DPQKFDPO...A..SRETLDH..	SIMY	[I][V][A][V][A][L]
Drosophila fucusphila (XP_017045444.1)	333	.DPQKFDPO...A..SRETLDH..	SIMY	[I][V][A][V][A][L]
Drosophila kikkawai	333	.DPQKFDPO...A..SRETLDH..	SIMY	[I][V][A][V][A][L]
Drosophila bipectinata	333	.DPQKFDPO...A..SRETLDH..	SIMY	[I][V][A][V][A][L]
Schistosoma haematobium	413	..AAGVFALALEAQNELTWRDQHLTVLT..	SKRNH..	LYDR	
Opisthorchis viverrini	598	SASVERFGMQYAE...KRRKTVPSLSSYLDFFVEQEVDFYKFPT			
Stylophora pistillata (PFX34234.1)	3707	...DNF.....	LKTIELDHNLALRY	[K][E][Q][S][L]V
Stylophora pistillata (XP_022787984.1)				
Nematostella vectensis				
Homo sapiens	251			
Branchiostoma	245	PETLQDF..NG..GHLLDKQDV	AIKRFPFA	
Branchiostoma belcheri	270	PETLQDF..NG..GHLLDKQDV	AIKRFPFA	
Mizuhopecten yessoensis (XP_021359564.1)	310	PKRLLDILTETDDVIL	HHQD	AIKRFPC
Mizuhopecten yessoensis (XP_021359565.1)	301	PKRLLDILTETDDVIL	HHQD	AIKRFPC
Mizuhopecten yessoensis (XP_021359566.1)	298	PKRLLDILTETDDVIL	HHQD	AIKRFPC
Ostrea edulis				
Crassostrea virginica (XP_022312773.1)	285	PSTLLEDVLNSEEVIL	HNQD	AIKSYPC
Crassostrea virginica (XP_022319601.1)	285	PSTLLEDVLNSEEVIL	HNQD	AIKSYPC
Crassostrea gigas (XP_011438844.1)	279	PSTFFQELQTSENLIL	HHQD	AIKSYPC
Crassostrea gigas (EKC41910.1)	300	PSTFFQELQTSENLIL	HHQD	AIKSYPC
Octopus bimaculoides	312	PDLLLEVALNMQHYIL	HNQD	AIKRFPA
Biomphalaria glabrata	341	PEMLKENVNPFFLL	HHQD	AIKRFPA
Lottia gigantea	315	PEHFLNHAGKKEFIL	HHQD	AIKRFPA
Pomacea canaliculata	302	PESFLAAHSSNKHFTL	HSQDV	AIKRYPA



APÊNDICE F – Árvore filogenética da CAD de metazoários

Arquivo externo

Árvore filogenética original da CAD de metazoários (relativo à Figura 4) inferida a partir do alinhamento de 72 sequências da CAD, presentes em 34 espécies de Metazoa. As sequências foram obtidas a partir das bases de dados eggNOG e Genbank. O alinhamento foi realizado com o alinhador Clustal Omega e analisado com o pacote Pvcust.

Apêndice G – Tabela de valores estatísticos 2
Árvore da CAD nos metazoários
Arquivo externo

Tabela contendo os valores de bootstrap (BP) e *approximately unbiased* (AU) calculados pelo pacote Pvcust, executado no ambiente R para a árvore filogenética da CAD nos metazoários (Figura 4).

APÊNDICE H – Árvore filogenética da Ccl nos três domínios

Arquivo externo

Árvore filogenética original da Ccl nos três domínios (relativo à figura 6). As sequências foram alinhadas com o alinhador Clustal Omega e os valores de confiabilidade foram gerados com o pacote Pvelust.

Apêndice I – Tabela de valores estatísticos 3
Árvore da Ccl nos três domínios
Arquivo externo

Tabela contendo os valores de bootstrap (BP) e *approximately unbiased* (AU) calculados pelo pacote Pvcust, executado no ambiente R para a árvore filogenética da Ccl nos três domínios (Figura 6).